technische universität
dortmund

Strangeness production in *p*-He and *p*-Ne collisions at $\sqrt{s_{\mathrm{NN}}} = 110$ GeV using the LHCb fixed-target mode

Dissertation zur Erlangung des akademischen Grades

Dr. rer. nat.

vorgelegt von

Lars Kolk

geboren am 28.10.1996 in Hamm

Fakultät Physik
Technische Universität Dortmund

Dortmund, Februar 2026

Der Fakultät Physik der Technischen Universität Dortmund zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften vorgelegte Dissertation.

Gutachter:
Prof. Dr. Johannes Albrecht
Prof. Dr. Christian Glaser

Vorsitzende der Prüfungskommission:
Prof. Dr. Julia Katharina Vogel

Vertreter der wissenschaftlichen Mitarbeiter*innen:
Dr. Gerald Schmidt

Datum des Einreichens der Arbeit:
13.02.2026

Datum der mündlichen Prüfung:
30.04.2026

Abstract

A measurement of multiplicity-dependent strangeness production in fixed-target collisions is conducted using data from the LHCb experiment at CERN. The measurement is motivated by observations of strangeness enhancement, the increase of strangeness production as a function of multiplicity, in various collision systems reported by the ALICE collaboration. Additionally, in the context of astroparticle physics, the presence of strangeness enhancement is of interest as it could explain the muon excess observed in extensive air showers, known as the Muon Puzzle. The analysis is performed on data collected with the fixed-target configuration of the LHCb experiment called SMOG, employing helium and neon as targets at a nucleon-nucleon centre-of-mass energy of 110 GeV. Differential cross-section ratios of the single strange particles K_S^0, Λ, and $\overline{\Lambda}$ to charged pions are measured in intervals of rapidity in the centre-of-mass frame and transverse momentum; for the study of strangeness enhancement, these ratios are additionally provided in intervals of multiplicity. The resulting cross-section ratios are compared with predictions of state-of-the-art hadronic event generators. None of the generators provide a consistent description of the data within experimental uncertainties. No significant hints of multiplicity dependent strangeness production are observed.

Kurzfassung

In dieser Arbeit wird eine Messung der multiplizitätsabhängigen Strangeness-Produktion in Fixed-Target-Kollisionen mit Daten des LHCb-Experiments am CERN vorgestellt. Die Messung ist motiviert durch Beobachtungen von erhöhter Strangeness-Produktion als Funktion der Multiplizität, genannt Strangeness Enhancement, die von der ALICE-Kollaboration in verschiedenen Kollisionssystemen gemacht wurden. Im Kontext der Astroteilchenphysik ist zudem das Auftreten von Strangeness Enhancement von Interesse, da es den als Myon-Rätsel bekannten gemessenen Myon-Überschuss in Luftschauern erklären könnte. Für die vorliegende Analyse wurden Proton-Helium- und Proton-Neon-Datensätze verwendet, die mithilfe des SMOG-Systems des LHCb-Experiments bei einer Nukleon-Nukleon-Schwerpunktsenergie von 110 GeV aufgezeichnet wurden. Gemessen wird das Verhältnis zwischen dem differenziellen Wirkungsquerschnitt von Teilchen mit Strangeness-Quantenzahl (K_S^0, Λ und $\overline{\Lambda}$) und dem differenziellen Wirkungsquerschnitt von geladenen Pionen. Die Messung erfolgt in Intervallen der Rapidität im Schwerpunktsystem sowie des Transversalimpulses. Zur Untersuchung der multiplizitätsabhängigen Strangeness-Produktion werden die Verhältnisse zusätzlich in Intervallen der Multiplizität gemessen. Die resultierenden Wirkungsquerschnittsverhältnisse werden mit den Vorhersagen gegenwärtiger Modelle hadronischer Wechselwirkungen verglichen. Keines der Modelle liefert dabei eine konsistente Beschreibung der Daten innerhalb der experimentellen Unsicherheiten. Es kann keine signifikante multiplizitätsabhängige Strangeness-Produktion beobachtet werden.

Contents

Appendices

1 Introduction

High-energy cosmic rays that enter Earth's atmosphere and interact with air nuclei produce secondary particles. These secondary particles can then interact with air nuclei again, thus triggering cascades of secondary particles, which are called *air showers*. Before the era of particle accelerators, measuring secondary particles produced in air showers was the only means of studying energetic particles [1]. First direct observations of cosmic rays were performed in 1927–1928 using cloud chambers in magnetic fields [2], rendering the trajectories of ionized particles visible. Only five years later, the use of cloud chambers led to the discovery of the first antiparticle, the positron [3, 4], for which the Nobel Prize in physics was awarded in 1936 [5]. In the same year, the discovery of the muon [6] followed, which was not predicted by any theoretical framework of the time. In the late 1940s and early 1950s, experiments using cloud chambers and photographic emulsion in high altitudes revealed two new types of neutral particles that decay into two opposite charged tracks; due to their characteristic V-shape pattern, they were coined *V-particles* and later identified as neutral kaons (K^0) and lambdas (Λ) [7–10]. During the same time, charged pions (π^+, π^-) [11, 12] were discovered in cosmic rays as well. The additional data collected at the Cosmotron [13], an early particle accelerator, made a deeper understanding of the V-particles possible. Their relatively long life-times hinted at a new quantum number [14–16], which was later called *strangeness* [17] and is now understood as the presence of a valence strange quark in the particle.

As accelerator technology advanced, laboratory accelerators soon replaced cosmic rays as the primary means of studying elementary particles, as they provided clean, controllable settings, allowing precise measurements. An era of accelerator physics began, leading to a series of newly discovered particles, see e.g. Ref. [18–20], and the establishment of a theory describing the interactions of fundamental particles and forces, the Standard Model (SM) of particle physics [21–26]. The description of the SM in its current form was finalized in the 1970s [27]. Since then, the SM was experimentally validated over the following decades. Precision measurements at high energy collider experiments, most notably at the Large Hadron Collider (LHC), operated by the European Organization for Nuclear Research (CERN), have played a central role in its experimental validation, see e.g. Ref. [28, 29].

In the preceding two paragraphs, the emergence of particle physics as a distinct discipline, separate from astroparticle physics, was outlined. However, the interplay between the two disciplines continues, and additional challenges arise when describing proton-nuclei interactions, as present in air showers. This is because these processes are dominated by regimes in which the strong interaction within the SM cannot be calculated from first principles. Thus, the description of hadronic interactions, i.e. interactions between hadrons governed by the strong, nuclear force, relies on phenomenological models. These models are implemented in Monte Carlo event generators, which are essential tools for the simulation of air showers. In modern day cosmic ray experiments, air showers can be described using two main features, the number of produced muons and the shower maximum X_{max}, which is the atmospheric depth at which the number of produced particles reached its maximum. There is, however, a long-standing discrepancy between the simulated and observed muon production in air showers, which implies missing physics in current state-of-the-art models of hadronic interactions. This discrepancy is called the Muon Puzzle [30]. A potential solution to the Muon Puzzle could be *strangeness enhancement*, which refers to the increased production of hadrons containing valence strange quarks as a function of energy-density. This effect was first only observed in heavy nuclei collisions [31–33], but was also observed in proton-proton and proton-lead collisions by the ALICE collaboration [31] at the LHC.

The short-coming of the event generators also makes it evident that additional data from proton-nuclei collisions are needed for their improvement or *tuning*. The LHCb experiment at the LHC offers a unique setting to study hadronic interactions. Equipped with the System for Measuring the Overlap with Gas (SMOG), the noble gases helium, neon and argon can be injected into the experiment, allowing it to be operated in a fixed-target mode. Using SMOG, the conditions occurring during air showers can be mimicked, providing important insights for the tuning of hadronic interaction models. Of particular interest would be measurements of proton-oxygen or proton-nitrogen collisions, as they are ideal proxies for air-showers. As proton-oxygen data only became recently available, the first step was to bracket oxygen using helium and neon.

In this thesis, the prompt production ratios of the lightest, neutral, primary strange particles—K^0_S, Λ, and $\overline{\Lambda}$, collectively referred to as V^0—to the lightest possible charged mesons, π^+ and π^-, are studied in proton-helium and proton-neon collisions at a nucleon-nucleon centre-of-mass energy of $\sqrt{s_{\text{NN}}} = 110\,\text{GeV}$. For this, several analysis steps have been performed that use the data gained from the LHCb experiment to separate particles of interest from background, thus allowing for the determination of differential cross-section ratios. A key feature of this thesis is the study of the multiplicity dependence of these ratios, which is closely related to the energy density of the system, thus giving insights on strangeness enhancement.

The author's contribution The main work of the author had a strong focus on data analysis and concluded in the measurement presented in Chap. 4. The presented measurement was performed within the Ion Fixed-Target (IFT) working group of the LHCb experiment and would not have been possible without the personnel and associated members of the experiment. The presented data analysis was developed in close collaboration with Dr. habil. Hans Dembinski, Prof. Dr. Michael Schmelling and Dr. Felix Riehn, who offered valuable scientific advice. The author is, however, the sole developer of this analysis and thus fully responsible for its contents. He was also fully responsible for the development of the underlying analysis code and internal documentation within the working group. By the time this thesis was submitted, the data analysis is undergoing the internal review of the IFT working group. Both the data and analysis code are centrally preserved.

The author also contributed to the preservation of several LHCb analyses within the RIVET and HEPData frameworks. In this context, the author also developed a guide on how to preserve analyses using HEPData and RIVET, which is published on the LHCb website. By developing dedicated online (HLT2) trigger lines, the author also established the basis for the selection of multistrange particle, i.e. particles containing multiple valence strange quarks, in the upgraded SMOG2 system, thus enabling future measurements. These contributions are presented in Chap. 5.

Outline This thesis is structured as follows: First, the theoretical foundation and context for the performed work is outlined in Chap. 2. Following this, the experimental setup, the LHCb experiment and its SMOG system, are introduced in Chap. 3. In Chap. 4, the performed measurement is presented. The results are then placed in the broader experimental context in Chap. 5, where there are compared to the measurements performed by the ALICE collaboration. In this chapter, additional contribution of the author within this context are discussed as well. Finally, a summary and outlook are provided in Chap. 6.

2 Context and foundations

In this chapter, the theoretical foundations and experimental context of this thesis are discussed. An overview of the SM is provided in Sec. 2.1. In Sec. 2.2, an introduction to hadronic interactions in air showers is given. This includes air shower measurements, hadronic event generators in air showers, and an introduction to the Muon Puzzle in air showers. Finally, possible approaches to solving the Muon Puzzle will be explored in Sec. 2.3.

2.1 The Standard Model of particle physics

The SM is a relativistic quantum field theory describing the nature of elementary particles and their interactions via the electromagnetic, weak and strong forces [21–26]. In the SM, elementary particles are divided into two categories; fermions with half-integer spin and bosons with integer spin. Fermions are particles that follow Fermi-Dirac statistics and are categorized in three generations of quarks and leptons, based on whether they participate in strong nuclear interactions. They make up the ordinary, known matter in the universe. For every fermion there exists also an antifermion with identical mass and spin, but inverted quantum numbers. Bosons, on the other hand, follow Bose-Einstein statistics and consist of gauge bosons with spin 1, which are associated with force mediators, and a scalar boson with spin 0, the Higgs boson. An overview of the elementary particles within the SM is given in Fig. 2.1.

There are two types of quarks: the up-type up (u), charm (c) and top (t) quarks, which carry an electromagnetic charge of $Q = +^2\!/_3\,\mathrm{e}$, and the down-type down (d), strange (s) and bottom/beauty (b) quarks with $Q = -^1\!/_3\,\mathrm{e}$. The first generation contains the lightest up- and down-type quarks, while the quarks in the second and third generations are progressively heavier. The s quark is notable here, as it is heavier than the u and d quarks, whilst being approximately 13.5 times lighter than the next heaviest quark, the c quark. This results in interesting properties, which will be discussed in more detail in Sec. 2.3. In addition to the electromagnetic charge, quarks carry one of three different *colour* charges, often referred to as *red*, *green* and *blue*; antiquarks carry the corresponding anticolour charges *anti-red*, *anti-green* and *anti-blue*.

Similarly to quarks, there are two types of leptons: the charged leptons with $Q = -1$, the electron (e^-), muon (μ) and tau (τ), and three neutrinos without electromagnetic charge, the electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ). The first generation of leptons contains the electron and electron neutrino, with the masses of the charged leptons progressively rising in the second and third generation. Although

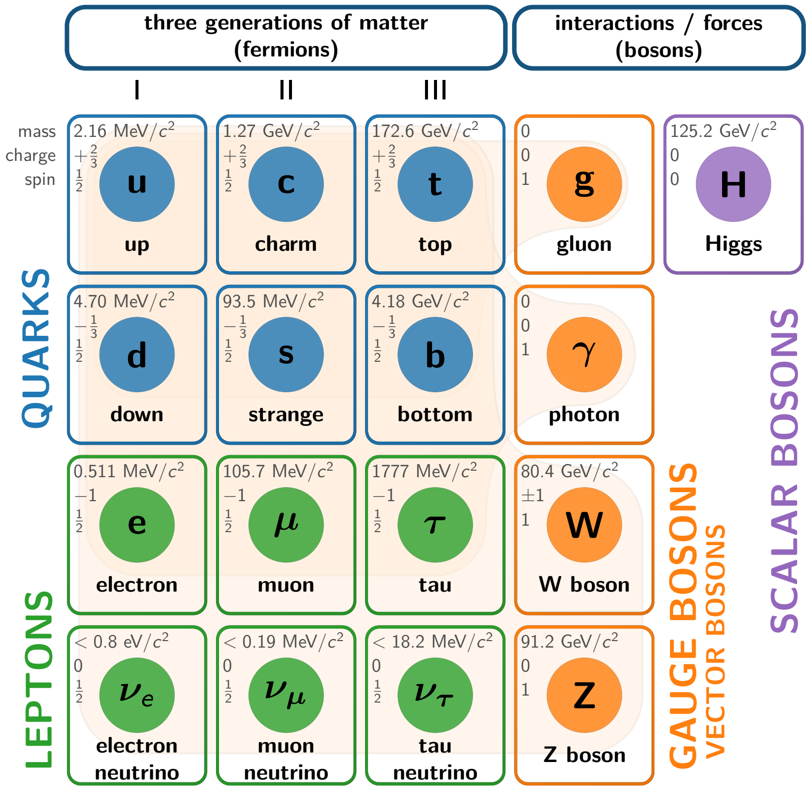


Figure 2.1 – Overview of the elementary fermions and bosons within the SM. Figure reconstructed from Ref. [34, 35], data taken from Ref. [36].

all neutrinos are assumed to be massless in the SM, neutrino oscillations have been observed [37], indicating that at least two neutrinos are massive.

The SM is a gauge theory described by the symmetry group $\mathrm{SU}(3)_\mathrm{C} \times \mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y}$. The $\mathrm{SU}(3)_\mathrm{C}$ symmetry group describes the strong force, where the label C denotes colour charge. The theoretical description of the strong interaction is also known as quantum chromodynamics (QCD). The gluons correspond to the generators of the group; they couple to colour charge while also carrying a colour- and anticolour-charge themselves. The gluon itself cannot be colour neutral, resulting in a set of eight linearly independent gluons, described by the Gell-Mann matrices [38]. A key characteristic of the strong interaction is that colour-charged particles must form colour neutral states, a phenomenon known as *confinement*. As a result, gluons cannot exist freely and quarks are forced to form bound, colour-neutral states with other (anti)quarks, called *hadrons*. The properties of hadrons are determined by *valence quarks*, which are bound together by gluons. Inside these hadrons, gluons can temporarily form quark-antiquark pairs called *sea quarks*. The most common hadrons are mesons and baryons. Mesons are made up of one valence quark and one valence antiquark with opposite colour charge, while (anti)baryons are made up of three valence (anti)quarks with three different (anti)colours. Exotic hadrons made up of four or more quarks/antiquarks have also been observed in the past few years [39–41]. Another key characteristic of the strong interaction is the running coupling constant α_s; at large momentum transfer, the coupling constant becomes small. As a result, quarks and gluons can be treated as free particles, a phenomenon known as *asymptotic freedom*. This regime is referred to as *hard QCD* or *perturbative QCD*, where perturbative refers to the mathematical method of expanding interaction terms in a power series to avoid divergence. This technique can only be used when the coupling constant is small. However, at low momentum transfer, the coupling constant becomes large; such processes cannot be calculated by first principles, thus creating the need for phenomenological models, which will be further discussed in Sec. 2.2.2. This regime is known as *Soft-QCD* or *non-perturbative QCD*.

The symmetry group $\mathrm{SU(2)_L} \times \mathrm{U(1)}_Y$ describes the electroweak interaction, the unification of the electromagnetic and weak interaction. This symmetry is however broken to $\mathrm{U(1)_{em}}$ via spontaneous symmetry breaking (SSB), which results from the excitement of the Higgs field. Before SSB, the $\mathrm{SU(2)_L}$ symmetry group contains three massless gauge bosons W_1, W_2, and W_3, which are the generators of the group and couple to particles with left-handed chirality (L). Massive particles always have a left- and right-handed chirality component, while the chirality of massless particles is identical to their helicity, the projection of the spin onto their momentum. The $\mathrm{U(1)}_Y$ symmetry group on the other hand contains the massless gauge boson B, corresponding to its generator. It is coupling to the hypercharge Y, which is defined via $Q = T_3 + {}^1\!/_2 Y$, where T_3 is the third component of the weak isospin. By definition, left-handed fermions have a spin of $T_3 = \pm {}^1\!/_2$ and form isospin doublets. Both quarks and leptons form one doublet within each generation: the up- and down-type quarks constitute one doublet, while the neutrino with its respective charged lepton form another, with the upper component being $T_3 = +{}^1\!/_2$ and the lower component being $T_3 = -{}^1\!/_2$.

After undergoing SSB, the three massive bosons and a massless photon emerge. The $W^\pm$ bosons emerge from the W_1 and W_2 bosons

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2) \tag{2.1}$$

and couple to the weak isospin. A unique property of the $W^\pm$ bosons is that they are able to change quark flavour. The photon and Z boson come from mixing of the B and W_3 bosons

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix}. \tag{2.2}$$

Here θ_W is the Weinberg angle, defined via the mass-ratio of Z and $W^\pm$ bosons:

$$\cos\theta_W = \frac{m_{W^\pm}}{m_Z}. \tag{2.3}$$

The resulting photon is massless and couples to electromagnetic charge, while the emerging Z boson is massive and couples to a combination of charge and weak isospin.

The final boson of the SM is the scalar Higgs boson with spin 0, which corresponds to an excitement of the Higgs field. The resulting SSB also gives rise to the fermion masses. Its discovery in 2012 by both the ATLAS and CMS experiments marked the final major addition to the SM [28, 29].

Despite its success, the SM is not able to describe several observed phenomena. As mentioned above, the neutrinos are massless in the SM, which is incompatible with the observed neutrino oscillations. Other open questions include the description of gravity as well as the description of dark matter and energy.

However, even within the framework of the SM, challenges arise. In particular the application of QCD in the description of air showers poses a challenge, as these processes are dominated by Soft-QCD. The resulting problems become evident when measuring the number of muons produced in air showers, which will be explored in the next section.

2.2 Hadronic interactions in air showers

Cosmic rays are fully ionized nuclei travelling through space, primary consisting of protons and elements up to iron, with a negligible proportion of heavier nuclei. The origins of cosmic rays are not yet fully understood. However, it is known that the vast majority of cosmic rays come from outside the solar system, with the highest-energy cosmic rays being of extragalactic origin [1, 30]. As illustrated in Fig. 2.2, their energies span 11 orders of magnitude with the flux of observable particles decreasing as a function of energy. Most notable are the transition regions at 3×10^6 GeV and at 3×10^9 GeV, commonly referred to as *knee* and *ankle*. While their exact origins are unknown, the knee is often associated with the end of the spectrum of galactic sources [1], while the ankle is often associated with the beginning of the spectrum of extragalactic sources [1].

Depending on their energy, cosmic rays can either be measured directly or indirectly. Up to energies of 10^5 GeV, space based experiments like AMS-02 [42], HEAO-3 [43] and CREAM [44] are used to measure cosmic rays directly. At higher energies, the flux of cosmic rays becomes so low that direct detection above the atmosphere is no longer feasible. Instead, they must be studied indirectly. High-energy cosmic rays that enter Earth's atmosphere interact with nuclei in the air, mainly consisting of oxygen and nitrogen, and produce secondary particles. These secondary particles can then interact with nuclei in the air again, thus triggering cascades known as air showers. These can be measured by ground-based experiments covering large areas, like the Pierre-Auger-Observatory [45] and Telescope Array [46], compensating for the low flux. The measurements of air showers will be covered in more detail in Sec. 2.2.1.

Air showers can be characterized by two main features, the number of produced muons and the shower maximum X_{max}, which is the atmospheric depth at which the number of produced particles reaches its maximum. Since there is no cosmic-ray source with a well-known composition, the interpretation of air showers relies on simulation [30]. As the propagation of air showers is dominated by Soft-QCD,

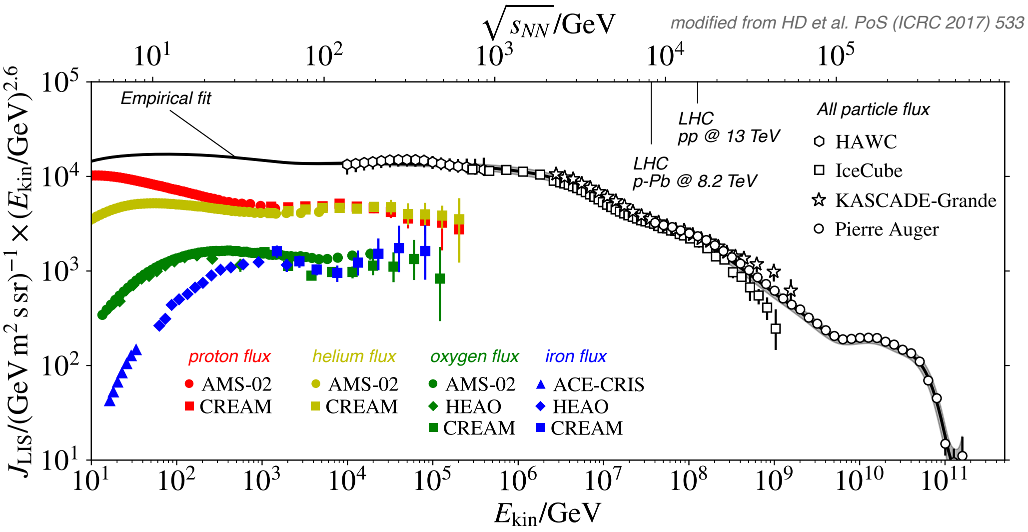


Figure 2.2 – Cosmic ray flux as a function of kinematic energy, where the flux is scaled by $E_{\mathrm{kin}}^{2.6}$, from Ref. [30]. The coloured markers show the fluxes of individual energies as measured by balloon- and satellite born experiments, while the white markers show the total flux.

phenomenological models are employed. While these models describe many features of air showers successfully, there is a long-standing discrepancy between simulated and observed muon production in air showers, which implies missing physics in current state-of-the-art models of hadronic interactions. This discrepancy is called the Muon Puzzle [30]. In the following sections, measurements of air showers and an overview of event generators in hadronic interactions are outlined.

2.2.1 Air shower measurements

Extensive air showers were first detected in the late 1930s by B. Rossi, W. Schmeiser, G. Bothe, W. Kolhörster and P. Auger [47]. The discovery was achieved using the newly developed *coincidence method*, in which multiple detectors could be placed several hundred of metres apart from each other, effectively linking them. When these detectors then registered near-simultaneous signals, it demonstrated that they must stem from a single high-energy cosmic ray. Although several groups contributed to this discovery, credit is usually given to Pierre Auger, as he provided the first clear quantitative evidence for correlated particle detection over large distances [48]. A complete historical review is given in Ref. [47]. Today, the world's largest cosmic ray detector, the Pierre Auger Observatory, is named after him and will be used as an example to explain how modern-day experiments perform air shower measurements.

A schematic view of an extensive air shower is presented in Fig. 2.3. As illustrated, extensive air showers are made up of a hadronic, electromagnetic and muonic component [49]. When a cosmic ray enters the atmosphere, it triggers a hadronic cascade where the average energy of the produced secondaries decreases exponentially with the number of interactions. When the hadrons produced in this cascade decay, they feed into the electromagnetic and muonic components of the shower. Neutral pions predominantly decay into photon pairs. These photons then create electron-positron pairs, which, together with bremsstrahlung, lead to the development of an electromagnetic cascade. Kaons and charged pions can decay into muons and muon neutrinos, creating a muonic component. The produced muons may also decay into electrons before they reach the ground, thus feeding into the electromagnetic component. All charged particles produced in the air shower can also ionize and excite air-molecules, which in turn emit fluorescence light when returning to the ground state [30]. The dense, central region of an air shower is referred to as the *shower core.*

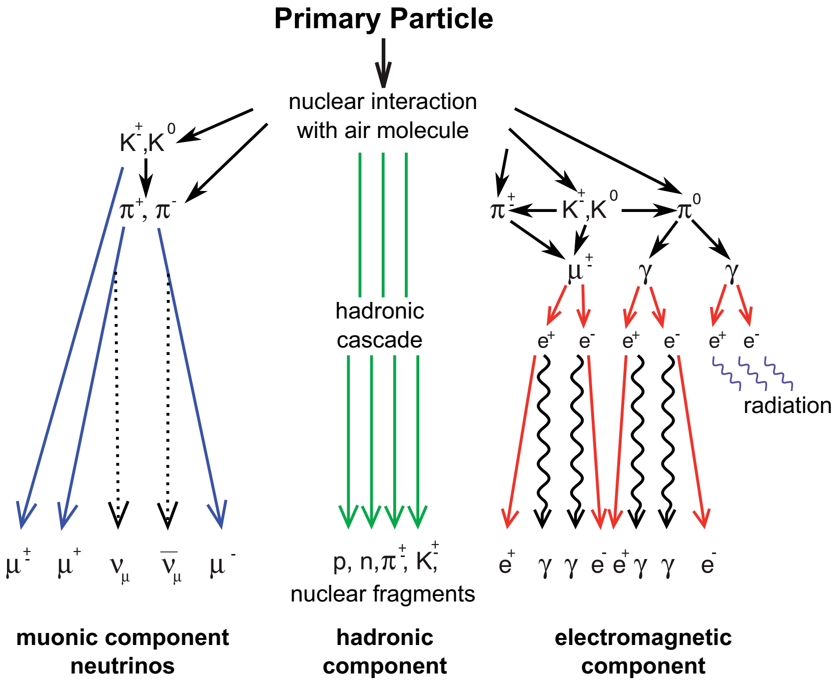
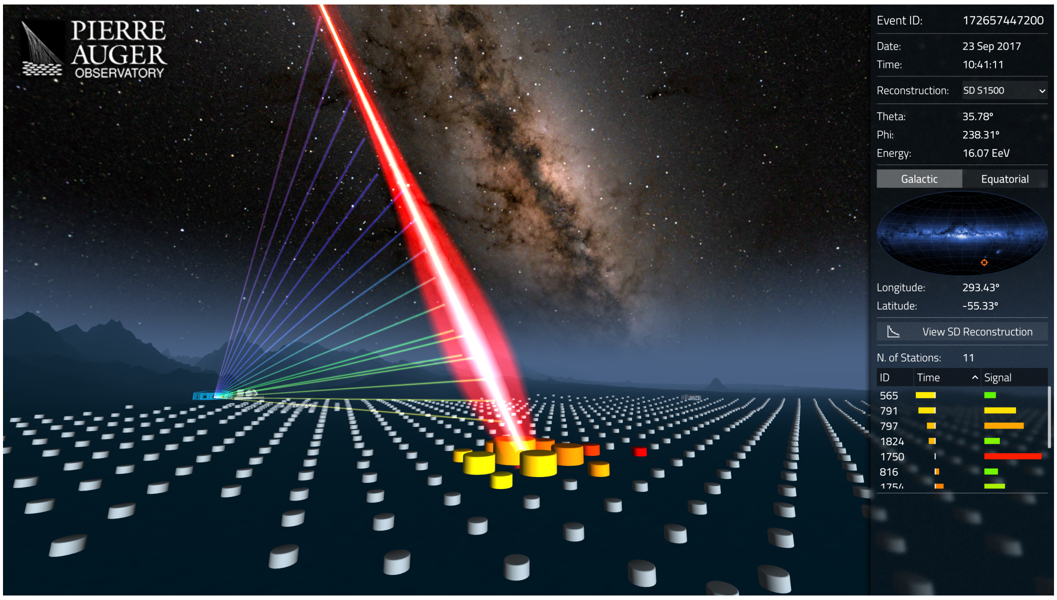


Figure 2.3 – Schematic view of an extensive air shower, from Ref. [49].

These fundamental properties of air showers form the basis for their observation; experiments like the Pierre Auger Observatory exploit them to reconstruct and measure the characteristics of the original cosmic rays. The Pierre Auger Observatory [45] is a *hybrid detector* completed in 2008 and located in Pampa Amarilla, Argentina; it uses an array of 1660 water Cherenkov particle detector stations spreading over 3000 km^2, together with 24 fluorescence telescopes to measure air showers. Each surface detector is filled with 12 tons of de-ionized water and is separated from its neighbours by 1.5 km. When a highly energetic charged particle enters a surface detector and travels faster in the water than light does, it produces Cherenkov light. This light is amplified by photomultiplier tubes and then converted into an electrical signal, enabling a measurement of the particles' properties. The number of measured signals per tank, together with the timing information, allows for as triangulation of the air shower and a determination of the shower core. The number of muons can be measured based on the signal arrival times in the surface detectors as well [30]. The fluorescence light produced in an air shower can be measured by the 24 fluorescence telescopes, allowing the observatory to track the development of the air shower and thus directly measure X_{max}. As these telescopes are highly sensitive, this is only possible in moonless nights. Measurements using information from both tanks and UV-telescopes are called hybrid measurements and have become an established standard in the field [30]. An example event display, illustrating the measurement of an air shower, is presented in Fig. 2.4.



Event ID: 172657447200
Date: 23 Sep 2017
Time: 10:41:11
Reconstruction: SD S1500
Theta: 35.78°
Phi: 238.31°
Energy: 16.07 EeV
Longitude: 293.43°
Latitude: -55.33°
N. of Stations: 11

Figure 2.4 – Example event display from the Pierre Auger Observatory of a hybrid event recorded by the both the ground array and the UV telescopes, from Ref. [50]. The red line represents the trajectory of the air shower. Each cylinder along the ground represents a surface-detector tank, where the colour indicates the time of signal detection and size indicates signal strength. The coloured lines from the blue building in the background, the fluorescence telescopes installed at Cerro Coihueco, indicate measurements of the UV light produced by the shower.

2.2.2 Event generators in hadronic interactions

The development of air showers is mainly driven by its hadronic component and thus hadron-ion collisions at low momentum transfer. Precise knowledge of QCD is therefore essential for the simulation and prediction of air shower features. There are however several challenges when describing hadronic interactions in cosmic-ray induced air showers. The initial collision of the cosmic ray and the nuclei in the air, which dictates how the shower will develop further, is highly asymmetric and complex due to the composite nature of colliding particles. Protons and nuclei are not elementary particles; a proton is a hadron composed of valence-quarks, gluons and sea-quarks, while nuclei are composed of protons and neutrons. These systems are complex and, as a result, it is not well understood how precisely these constituents collectively behave. Consistently describing different cosmic-ray types poses an additional challenge, as a cosmic ray can be any ionized nucleus, as established in Sec. 2.2.3. An additional challenge lies in the consistent description of both the production of the different secondary hadrons and their interactions with the nuclei. Moreover, the described processes typically occur at low momentum transfer, and thus fall in the regime of Soft-QCD, meaning that they cannot be calculated from first principles. As a result, phenomenological models must be employed to describe the development of cosmic-ray induced air showers.

The modelling of hadronic interactions generally consists of two parts, the description of the initial interactions of quarks and gluons, collectively referred to as partons, and the subsequent treatment of hadronisation processes and collective effects. A common ansatz is to describe parton interactions via exchanges of effective particles and fields, simplifying the complex interactions. The mathematical framework of such an ansatz is formulated as an *effective field theory* and the properties of hadronic interactions are typically described via Gribov-Regge theory [51, 52]. This theory introduces colourless particles called *Pomerons*, carrying the quantum numbers of the vacuum, to describe the slowly rising cross-sections in high energy hadron collisions. A major drawback of Gribov-Regge theory is that energy is only conserved at particle production level, but not for the calculation of cross-sections [53]. Current state-of-the-art event generators based on Gribov-Regge theory include `QGSJet-II-04` [54], `EPOS-LHC` [55] and `SIBYLL-2.3d` [56]. Although these generators differ in their underlying assumptions and modelling, they all come with a set of adjustable parameters chosen to reproduce particle- and astroparticle data. Hadronic interactions can also be described via *multiparton interaction*. In this approach, perturbative QCD is not only employed to describe hard processes, but also in the description of soft interactions by calculating multiple $2 \rightarrow 2$ scattering processes. The divergence of the total cross-section is then avoided by applying an energy-based cut-off. A current state-of-the-art model following this approach is `Pythia 8` [57].

The process of hadronisation is commonly described via *string-fragmentation*. Due to confinement, partons must exist in bound states. When two partons are separated, the gluon field between two colour charges forms a narrow flux tube, which is also referred to as string. With increasing distance, the potential between the two colour charges becomes large enough to create new quark-antiquark pairs from the vacuum, thus breaking the string. In the implementation of *standard string fragmentation*, known as *the Lund string* [58], each string is handled independently without considering spatial correlations or interactions with other strings. However, in real hadronic interactions this idealized model may not show the full picture. At high energies, many partons and strings are produced, which can radiate or interact with each other before hadronising, leading to collective effects. In the models listed above, several approaches are realized, which are summarized in the following. One approach is the implementation of *colour-reconnection* [59], in which colour-connections can be rearranged before hadronisation, thus minimizing string energy and increasing hadron, but not strange, production. Another approach is to consider interactions of strings. One possible mechanism is *string shoving* [60, 61], the repulsive interaction between two overlapping flux tubes, giving rise to transverse collective flows in high-density systems [62]. Another type of string-string interaction is *rope hadronisation* [63], in which dense regions of spacially overlapping strings form a rope with high energy-density. The rope then fragments into quark-antiquark pairs, leading to increased production strange particles and baryons. A similar mechanism is the thermal description of hadronisation in regions of high parton density [64]; these regions are treated as locally thermalized cores consisting of deconfined quarks and gluons. These cores evolve hydrodynamically and produce hadrons statistically until they freeze out.

Between the initial interaction and hadronisation, some generators simulate initial and final state radiation. In these processes, radiation—often in the form of gluons and photons—is emitted in the initial or final state of the interaction. A brief overview of the four generators and their approaches used to model hadronic interactions is given below.

EPOS-LHC The `EPOS-LHC` model is a special version of the `EPOS` [65] model, specifically developed to describe the data collected at the LHC. The models are based on a semi-hard Pomeron approach, meaning that they use both aspects from Gribov-Regge theory and perturbative QCD. This approach includes a mechanism analogous to initial and final state radiation, resulting in a broader hadron spectrum in rapididity. A characteristic feature of `EPOS` is that, unlike other Pomeron-based models, energy conservation is explicitly maintained in the calculation of cross-sections. In `EPOS-LHC`, hadronisation is modelled via standard string fragmentation at low energy density, and via thermal hadronisation at high energy density.

QGSJet-II-04 Similar to the `EPOS` models, `QGSJet-II-04` is based on a semi-hard Pomeron approach. It is however explicitly designed for air shower simulation and only physics deemed relevant for air showers is implemented. It has a very limited parameter set and the fewest parameters among the models presented here. In `QGSJet-II-04`, hadronic interactions are exclusively described via an early implementation of the Lund string model.

SIBYLL-2.3d `SIBYLL-2.3d` is another model explicitly designed for the modelling of general air shower features, where only physics deemed relevant for the air shower is implemented. It employs a *dual parton* [66] approach; processes in Soft-QCD are calculated based on Gribov-Regge theory, while processes in hard QCD are calculated based on perturbative calculations, where an energy dependent cut-off in low transverse momentum is applied. In `Sibyll-2.3d` hadronisation is modelled via standard string-fragmentation. The model is optimized for cosmic-ray physics and only the production of hadrons containing valence u, d, s and c quarks is considered.

Pythia 8 `Pythia 8` is a general purpose generator which is primarily used at the LHC to simulate hard and soft processes in proton-proton interactions. Unlike the other models listed here, it is not based on Gribov-Regge theory, but multiparton interactions instead. Alongside multiparton interactions, `Pythia` also simulates initial and final state radiation. The cross-section for semi-hard partonic scatterings are calculating via perturbative QCD. In doing so, a transverse-momentum threshold is applied that is regulated via a smooth damping factor. The calculations of cross-sections is however performed separately from particle production [67]. In the standard tune of `Pythia 8`, hadronisation is modelled via *standard string fragmentation*, taking no collective effects into account. However, `Pythia 8` includes also implementations of colour-reconnection, string-shoving and colour ropes, which can be enabled in specific tunes.

Even though these state-of-the-art generators offer detailed implementations of hadronic processes, their application in the description of the hadronic air shower component is currently the largest source of uncertainty [68, 69]. Furthermore, while the models successfully predict features such as the shower depths, discrepancies arise when comparing the number of simulated and observed muons, which will be discussed in the next section.

2.2.3 The Muon Puzzle in astroparticle physics

In 2015, the Pierre Auger Collaboration presented their first hybrid measurement of the mean muon number in cosmic-ray induced air showers [70, 71]. This measurement, following a nearly model-independent approach, revealed that the hadronic interaction models consistently predict fewer muons than observed. Following this observation, numerous experiments conducted additional measurements. While the deviations in the reported muon counts did not individually exceed a significance of 3σ, a meta-analysis taking all data into account revealed a significance of 8σ [72]. Even after the existing model parameters were adjusted—a process called *tuning*—the discrepancy persisted and became known as the *Muon Puzzle* [30]. Comparisons of the observed and simulated muon numbers are presented in Fig. 2.5. As illustrated there, `QGSJet-II-04`, `EPOS-LHC` and `SIBYLL-2.3d` consistently underestimate the muon count, both as a function of average shower depths and energy. For the latter, comparisons to proton and lead showers are provided as well. As established earlier, the majority of cosmic rays consist of protons, with heavier nuclei being significantly less likely. However, even when assuming exclusively iron showers, the muon count is still underestimated in the majority of intervals.

A key property of the Muon Puzzle is that the only modifications resolving these discrepancies would either violate constraints from data or lead to inconsistencies in the description of other air shower features. This indicates that physical effects may be missing in the description of hadronic interactions [30], which will be further explored in the next section.

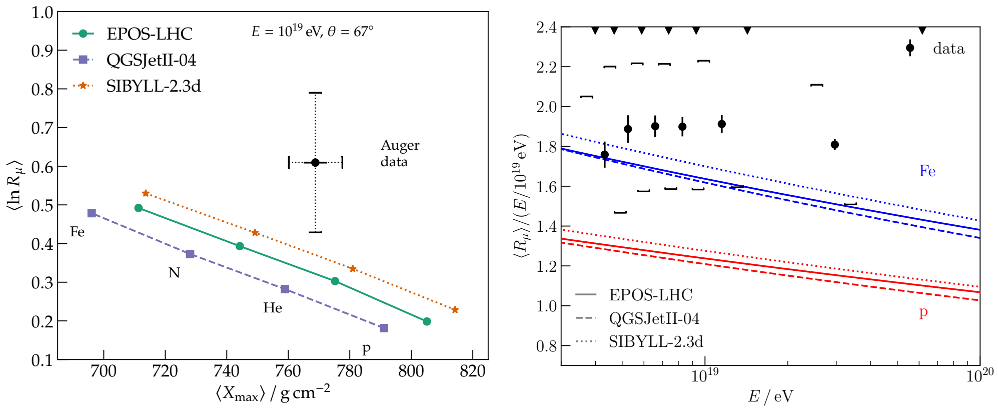


Figure 2.5 – Measured and simulated muon numbers, from Ref. [73]. In the left plot the average logarithmic muon count as a function of average shower depth is shown, while the average number of muons as a function of energy, together with predictions for proton and lead showers, is shown in the right plot.

2.3 Finding solutions to the Muon Puzzle

In this section, several approaches to solving the Muon Puzzle are discussed. First an introduction to strangeness enhancement, a mechanism that may be responsible for the surplus of muons produced in air showers, is given. This is followed by an outline on how centralized databases and standardized interfaces for data-simulation comparisons in particle and astroparticle physics can be used to reduce uncertainties in air shower data.

2.3.1 Strangeness enhancement

Simulations of extensive air showers have shown that relative reduction of the electromagnetic component in air showers could explain the muon excess in data. Extensive studies on how such a modification could affect air shower observables were performed in Ref. [74] by varying the hadron multiplicity and mean energy ratio,

$$R = \left\langle \frac{\mathrm{d}E_{\mathrm{EM}}}{\mathrm{d}\eta} \right\rangle \Bigg/ \left\langle \frac{\mathrm{d}E_{\mathrm{hadr.}}}{\mathrm{d}\eta} \right\rangle , \tag{2.4}$$

where E_{EM} and $E_{\mathrm{hadr.}}$ are the energies of the electromagnetic and hadronic component, and η is the pseudorapididy of all final state particles, except neutrinos. The effect on the simulated number of muons and shower-depth are visualized in Fig. 2.6. It is shown that a reduction of R would not significantly change $X_{\max}$ while increasing the number of muons enough to be compatible with the data taken by the Pierre Auger Observatory.

Therefore, a mechanism that reduces R could potentially solve the Muon Puzzle. In the initial reaction, there is a roughly equal production of π^+, π^- and π^0 mesons

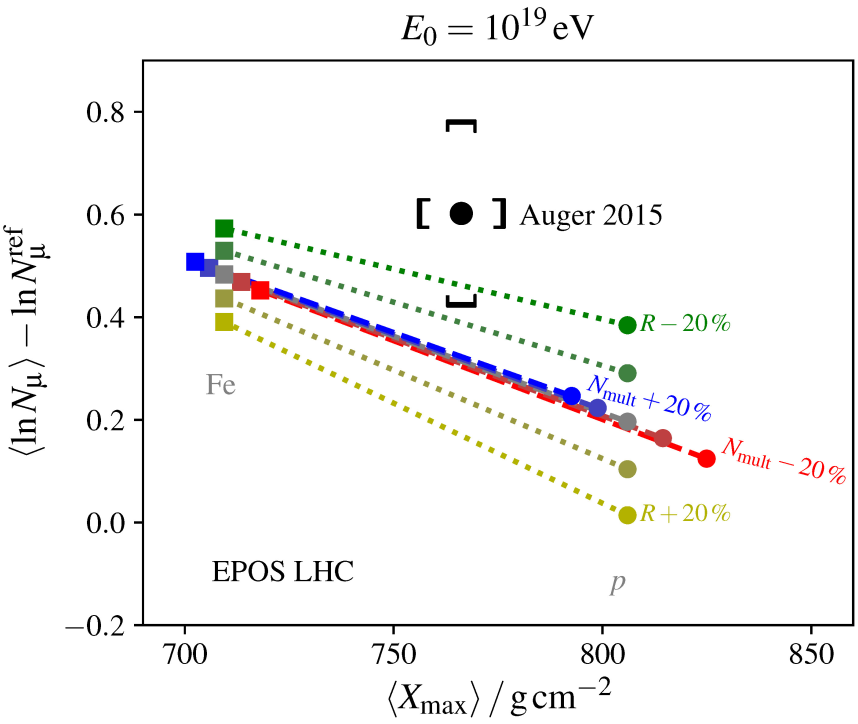


Figure 2.6 – Impact of varying the hadron multiplicity N_{mult} and energy ratio R in 13 TeV LHC collisions on `EPOS-LHC` model predictions, from Ref. [74]. The predictions are compared to the data collected at the Pierre Auger Observatory in 2015 [70]. The data are represented by a black point, with the uncertainties indicated by the end-caps. The model lines represent all values that can be obtained by any mixture of iron (top left) and protons (bottom right).

due to isospin symmetry. Due to this symmetry, the initial ratio of charged to neutral pions cannot be altered. One possibility to reduce R would be however an increased production of strange particles, i.e. particles containing valence strange quarks. With increased strangeness production, energy which could otherwise go into the production of pions would instead go into the production of kaons or other strange hadrons. Unlike neutral pions, kaons do not immediately decay into two photons, thus keeping energy in the hadronic channel longer. In addition, charged kaons decay predominantly into muons, thus directly contributing to the muon component.

An increased production of strange particles—known as *strangeness enhancement*—was initially only observed in nucleus-nucleus collisions [31–33]. It was linked to the emergence of a *quark-gluon plasma*, a state which nuclear matter can reach at high temperatures and energy densities in which quarks and gluons are no longer confined into colour-neutral hadrons. Recent findings of the ALICE collaboration show that strangeness enhancement is also observed in multiple systems of proton-nucleous and proton-proton collisions [31]. The ratios of different particle yields to charged pions, as found by the ALICE collaboration, are presented in Fig. 2.7. The ratios are shown as a function of average multiplicity at mid-rapididity, which is used as a proxy for the overall event activity and thus energy-density in the collision. It is observed that the ratios show a universal rise in strangeness production between different collision systems and energies. It is also observed that the enhancement is larger for the multi-strange particles Ξ^- and Ω^- than for the single-strange particles K^0_S, Λ and $\overline{\Lambda}$. A comparison with generator predictions is also given. Apart from `EPOS-LHC` and `Pythia 8`, the ALICE collaboration provides comparisons with `DIPSY` [75], in which colour ropes are used to model hadronisation.

The comparisons between the event generators and data show that `Pythia` does not describe the observed enhancement, indicating that there is a mechanism missing in its standard tune. `EPOS-LHC` and `DIPSY` show an enhancement with the same trend as the data. However, both require retuning to describe the data consistently. Furthermore, additional data is needed to decide which mechanism of strangeness enhancement may be realized in nature. In particular, measurements involving additional collision partners and centre-of-mass energies would be of great interest.

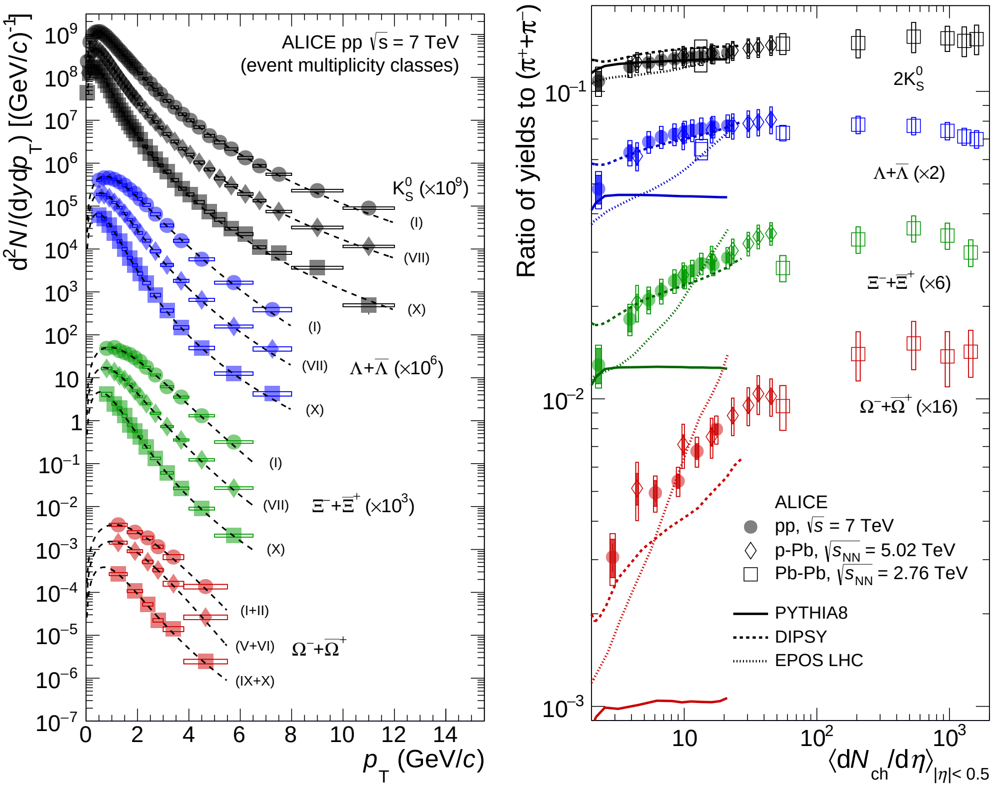


Figure 2.7 – Yields of strange particles for different collision systems and energies, as measured by the ALICE collaboration from Ref. [31]. The left plot shows the scaled particle yields of different multiplicity intervals, where the black dashed lines are fits describing the curves. The right plot shows the scaled ratio yields to charged pions, together with generator predictions.

2.3.2 Tuning of event generators

As outlined in Sec. 2.3.1, clarifying which mechanisms of strangeness enhancement is realised in nature will require more data. In addition, hadronic interactions remain the overall dominant source of uncertainty in the simulation of air showers overall. Reducing these uncertainties requires numerical results of experimental measurements to be made readily available, ideally in a machine-readable format.

In the particle physics community, a central open-access database is provided via the High-Energy Physics Database (HEPData) [76], which is currently hosted by CERN. Within HEPData, the numerical results of over 10 000 published measurements are stored as *records*, out of which approximately 2 000 stem from LHC experiments. Each record has a unique identifier linking it to its publication, and consists of a set of tables, in which the results are stored and briefly described. The tables can be downloaded in various, machine-readable formats.

However, providing the results in machine-readable format alone is not sufficient. For comparisons between data and generator predictions to be meaningful, these comparisons must involve identical selections, studied kinematic intervals, and definitions of observables. The latter is especially important, as different communities may use the same word for different concepts, or employ slightly different criteria. A standardized framework for the validation of the event generators is the Robust Independent Validation of Experiment and Theory (RIVET) toolkit [77]. Within RIVET, published measurements are preserved as *plugins*, each consisting of multiple files containing metadata, measurement results, selection code and plotting instructions. The measurement results are directly obtained from HEPData, thus creating a strong link between the two frameworks. The selection code is the main part of the plugin and written in the C++ programming language. It is used to process generated events from the event generators. This is done by iterating over the events, selecting processes of interest analogous to the measurement, and constructing the corresponding observables. In doing so, the metadata can also be accessed, ensuring that the generated events meet the measurement conditions, including centre-of-mass energies and collision partners. Using the defined plotting routine, RIVET provides visual comparisons between the measured and generated observables. While several analyses are already implemented within the RIVET framework, the RIVET team continues to maintain an extensive wishlist of analyses that have yet to be added, including many measurements from the LHC experiments. Contributions of the author within the RIVET and HEPData frameworks are summarized in Sec. 5.3.

While an equivalent to HEPData exists in the astroparticle community [78], a RIVET-like translator is still missing. Such a translator would be an important step towards a combined tune, using data from both particle and astroparticle physics. A roadmap on how such tunes could be achieved is given in Ref. [67]. While the tuning on existing data is an essential step towards the solution of the Muon Puzzle, additional measurements are still required, as discussed in Sec. 2.3.1. Therefore, a dedicated experiment capable of providing precise measurements in multiple collision systems is needed. The LHCb detector, which fulfils this these requirements, is presented in the subsequent chapter.

3 Experimental setup

In this chapter, an overview of the experimental environment is given. First, the LHC is introduced which provides the necessary, high-energy proton and ion beams. Next, the LHCb experiment allowing the measurement of particle collisions is presented. Finally, an overview of the SMOG System is given, which enables LHCb to be operated as fixed-target experiment.

3.1 The Large Hadron Collider

The LHC [79, 80] is the largest and most powerful particle accelerator build to date. It is a synchrotron spanning a circumference of 26.7 km, embedded into the broader CERN accelerator complex depicted in Fig. 3.1. The accelerator complex is located underground, near the Franco-Swiss border in the Geneva area and is designed to accelerate charged particles to energies of up to 7 TeV; by accelerating charged particles opposite directions and colliding them, the LHC can reach a design centre-of-mass energies of up to $\sqrt{s} = 14$ TeV. Since the LHC has started operation, there were several periods dedicated to data taking referred to as *Runs*, between which the LHC and the experiments located at the LHC have either been commissioned or upgraded. A short overview of the performed and planned Runs is given in Ref. [81, 82]. The measurement presented in this thesis was performed using data from Run 2, which started 2015 and ended 2018. During Run 2, the beams reached energies of up to 6.5 TeV, corresponding to a centre-of-mass collision of 13 TeV for colliding beams. Following Run 2, Run 3 started in 2022 and is now scheduled to end mid 2026. For Run 3, the beam energies were increased to up to 6.8 TeV.

The LHC does not accelerate individual particles, but packages of particles referred to as *bunches*. While the charged particles within a bunch are typically protons, the LHC is also capable of accelerating ions of other elements, as such lead, xenon, neon and oxygen. However, as protons are the main particle species accelerated in the LHC, the remainder of this section will focus exclusively on proton bunches.

A beam of protons is made up of 2808 bunches with a time spacing of 25 ns, where each bunch contains 1.2×10^{11} protons at the start of injection. These protons are obtained via the ionization of hydrogen-gas and initially accelerated using LINAC2, a linear accelerator, to an energy of 160 MeV. The Proton Synchrotron Booster subsequently increases their energies to 2 GeV before injecting them into the Proton Synchrotron, in which they are further accelerated to 26 GeV. From there, the protons enter the Super Proton Synchrotron, where they are accelerated to 450 GeV. Finally, the proton beams enter the LHC, where they are further accelerated until they reach their target energy. In the LHC, the two beams are then brought to collision at dedicated beam-crossing positions through an adjustment of the magnets and are kept in circulation until their intensity has decreased significantly. The data taking period of from a single injection is referred to as *fill*.

The interaction points are located in caverns capable of hosting specialized experiments, able to analyse the outcome of these collisions. The four main experiments of the LHC are ALICE, ATLAS, CMS and LHCb. The ALICE experiment [83, 84] specializes in the analysis of quark-gluon plasma, as present in heavy-ion collisions. The ATLAS [85, 86] and CMS [87, 88] experiment are general purpose detectors; their geometries are optimized to achieve broad coverage in pseudorapididy. They are best known for their discovery of the Higgs boson in 2012 [28, 29]. Finally, the LHCb experiment is specifically designed for the study of bottom- or charm-hadrons. Unlike the other experiments, it is not build symmetrical around the interaction point and focuses on the forwards-region in which bottom- and charm-hadrons are predominantly produced. Since Run 2, it has also been equipped with the SMOG system [89]. This system allows LHCb to be operated in a fixed-target mode, in which one incoming beam collides with stationary gas. A detailed description of the LHCb experiment is given in the following section.

3.2 The LHCb detector

The LHCb detector is a single-arm forwards spectrometer specifically designed to for precision measurements of processes involving bottom or charm hadrons. At LHC energies, bottom and charm hadrons are dominantly produced via gluon-gluon fusion processes [91], involving gluons from the partonic content of the incoming projectiles. The momenta carried by the gluons is likely to be asymmetric; one gluon carries large fraction of longitudinal momentum, whereas the other gluon only carries a small fraction of longitudinal momentum. As a result, the created particles experience a boost along the beam axis, causing bottom and charm hadrons to be predominantly created under small angles in forward and backward direction. Due to limited space in the cavern and due to cost reasons, the LHCb detector only covers the forward rapidity region $\eta_{\text{lab}} \in [2, 5]$, where η_{lab} is the pseudorapidity in the laboratory system. Pseudorapidity is a non-linear mapping of the angle θ between the particles momentum and beam axis and is defined as $\eta_{\text{lab}} \equiv -\ln\left[\tan\frac{\theta}{2}\right]$. It can be related to rapidity, which is an additive measure of velocity under Lorentz transformation and defined as $y \equiv \frac{1}{2}\ln\left(\frac{E+p_{\text{L}}}{E-p_{\text{L}}}\right)$, where p_{L} is the component of the momentum along the beam axis.

The LHCb detector is visualized in Fig. 3.2. It spans approximately 20 m in length, 10 m in height and 12 m in width. The detector consists of several components dedicated to Particle Identification (PID) and the reconstruction of particles trajectories and interaction vertices, which will be briefly summarized in the following. Components specifically relevant for PID, tracking and vertex reconstruction will be additionally highlighted in dedicated sections. Particles are brought to collisions inside the Vertex

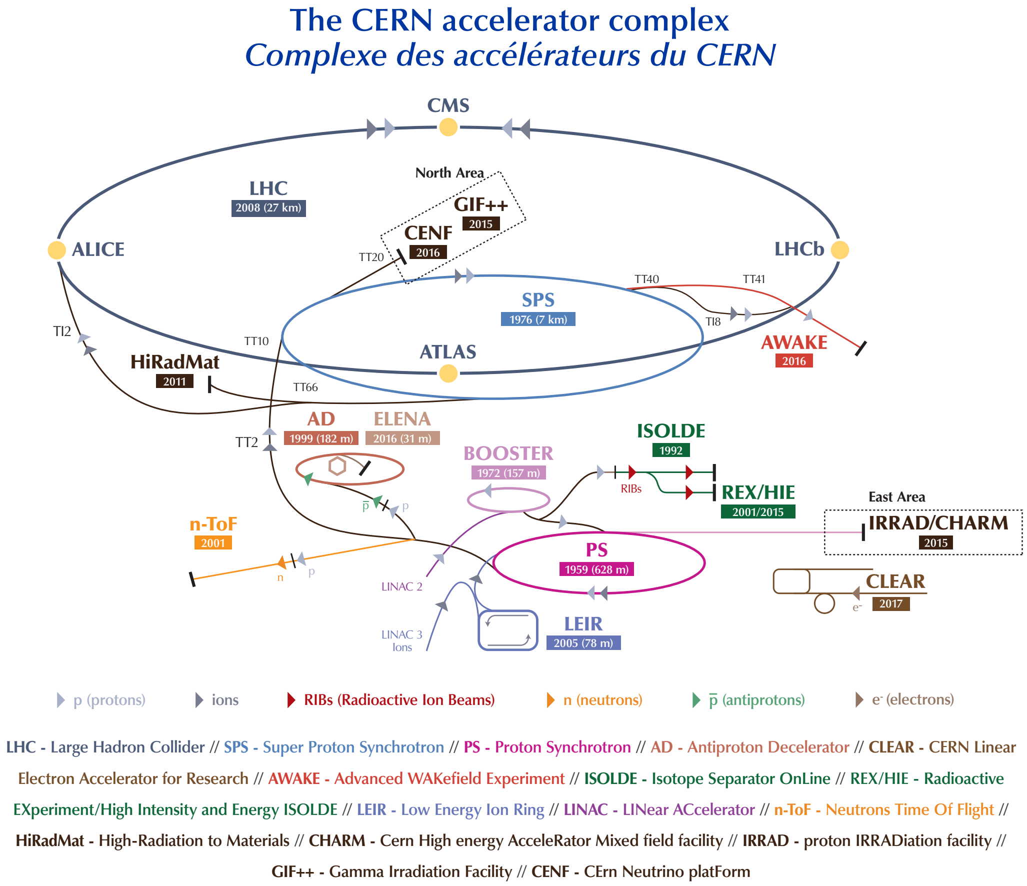


Figure 3.1 – CERN accelerator complex as of 2018. Figure taken from Ref. [90].

Locator (VELO), where the primary interaction point, the primary vertex, is reconstructed. Secondary vertices originate from the decay of short-lived particles and play a key role in particle reconstruction; the spatial distance between primary and secondary vertex gives important insights about the flight-distance—and hence lifetime—of undetected particles. Together with the Tracker Turicensis (TT), placed upstream of the dipole magnet, and the subsequent tracking stations T1, T2, and T3 placed downstream of the dipole magnet, the VELO plays a crucial part in the reconstruction of trajectories of charged particles. The dipole magnet plays an important role in bending the trajectories of charged particles, allowing the determination of charges and momenta. The Ring-Imaging Cherenkov (RICH) detectors, RICH1 and RICH2, are mainly used to determine the species of charges particles. The RICH1 detector is placed between VELO and TT, while the RICH2 detector is placed downstream of T3. The calorimeter system, placed downstream of RICH2, consisting of the scinillating-pad detector (SPD), preshower (PS) detector, electromagnetic calorimeter (ECAL) and hadronic calorimeter (HCAL), provides additional separation power. Lastly, the detector employs five muon chambers, labelled M1-M5, responsible for muon-identification. The chambers M2-M5 are placed downstream of the calorimeter system, while M1 is placed downstream of RICH2 and upstream the calorimeter system. The calorimeters, muon chambers and RICH detectors all serve as PID system.

The LHCb detector is also equipped with the SMOG system [89], which enables the injection of the noble gases helium, neon and argon into the VELO vessel. Combined with the forward geometry of the LHCb detector, this enables LHCb as the only one of the four major experiments to be operated as a fixed-target experiment.

During the majority of Run 2 operation, the beam parameters were configured to maintain a constant luminosity across the duration of the fill, a procedure known as *leveling*. This leads to constant luminosities across the duration of a fill, which the LHCb detector was designed to operate on. The average bunch crossing frequency is 30 MHz [92], corresponding to one collision every 25 ns. Storing this amount of data throughout the entire duration of Run 2 is neither feasible nor efficient. Therefore, a trigger system consisting of a hardware trigger (L0) and a subsequent two-staged software-trigger are employed. The L0 relies on information from both the calorimeter system and the muon chambers. For an event to pass the L0 trigger, it must exceed either the minimum transverse momentum, p_{T}, threshold in the muon chamber or deposit a minimum transverse energy threshold in the calorimeter system. An additional veto is applied on events with exceedingly high occupancy, and events in which multiple collisions are recorded. The latter is also known as *pile-up* events. The veto is based on the number of hits recorded in the SPD and PS detectors. The output-rate of the L0 trigger is 1 MHz. All events that pass the L0 trigger are then further processed by the software triggers, consisting of the HLT1 and HLT2 triggers. In this step, events are selected based on information gained from first their partial and then full reconstruction. While the exact criteria vary depending on the physics programme, the HLT1 trigger generally selects events based on transverse momenta and impact parameters of reconstructed particles. This is done using only information from the tracking stations. The HLT2 trigger on the other hand makes full use of the PID information and is typically used to filter for specific decay topologies. The selection of events using the software trigger is also referred to as *online* selection. A detailed summary of the design, performance and challenges of the trigger system is given in Ref. [93].

With the end of Run 2, the LHCb detector underwent a major upgrade, with the goal to increase the instantaneous luminosity by a factor of five. This invoked two main challenges: In order to deal with the increased data, the trigger system had to be upgraded; the hardware-trigger was removed and the software-triggers were adapted, resulting in an online-only trigger system. As a result, several detector components had to be upgraded or replaced in order to handle the increased luminosities and radiation damages, while also providing faster readouts of 40 MHz. An extensive description of the upgrade is given in Ref. [94]. During this time, the SMOG system was upgraded as well [95]. Details regarding the SMOG system are discussed in Sec. 3.2.3.

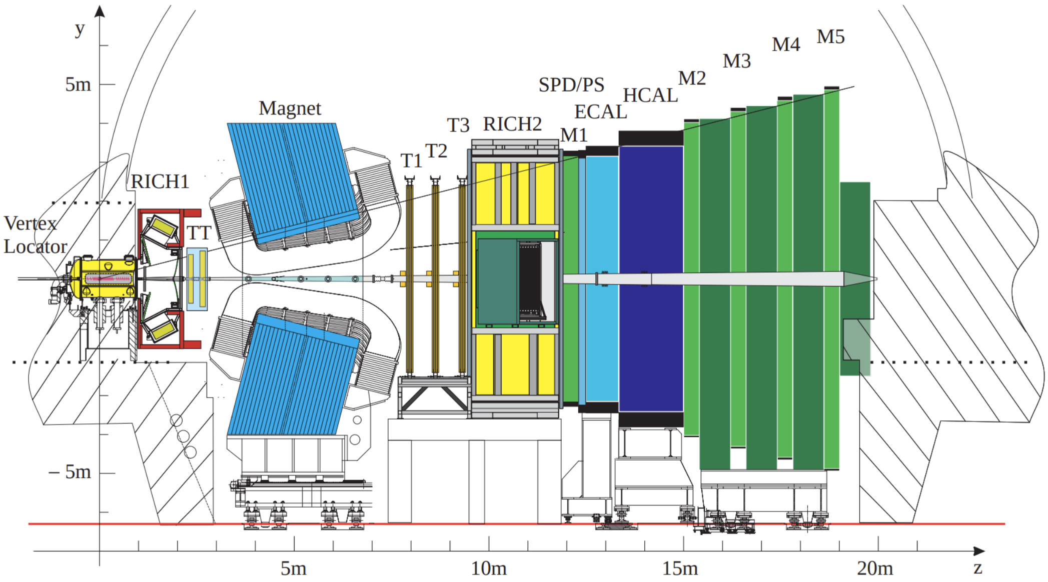


Figure 3.2 – Cross-section of the LHCb detector and its components, from Ref. [96]. The coordinate system is a right-handed Cartesian system with the origin in the centre of the nominal proton-proton interaction region; the x axis is horizontal and orthogonal to the beam-pipe, the y-axis is pointing towards the surface and the z-axis is pointing along the beam-pipe.

3.2.1 Track and vertex reconstruction

The goal of the tracking system is the reconstruction of particle trajectories and vertices, enabling the determination of their momenta and charges. The reconstruction of vertices is especially important, as it allows the spatial reconstruction of decays of particles. The VELO [97] is one of the most crucial components of the detector and thus the tracking system. It is a silicon microchip detector around and downstream the proton-proton interaction region, arranged in two halves along the beam axis. Each half consists of 21 semicircular modules, each providing measurements as a function of radius, r, and azimuthal angle, ϕ, around the beam axis. When no particle collisions take place, the two halves are retracted to protect them from potential damage. This is essential as there is no beam-pipe around the interaction point, exposing the VELO which is only protected by a thin aluminium-alloy foil. Once stable conditions are reached, the VELO is moved into its operating position, brought to a minimum distance of 7 mm around the beam axis. At this proximity, the VELO provides nearly full coverage of the LHCb phase space and only particles with either very low or high pseudorapidity are not covered by it. The trajectory of charged particles are reconstructed from their hits—the deposition of energy in the sensors via the creation of electron-hole pairs—in successive detector layers. These hits are then used by dedicated algorithms to reconstruct the trajectory, referred to as *track*. Inside the VELO, they can be reconstructed as straight lines, as the effect of the subsequent dipole magnet are negligible. Within the geometric and kinematic acceptance of LHCb, the VELO provides a track-finding efficiency of at least 98 %. In addition to its high track-finding efficiency, the VELO provides high spatial resolution of up to 4 µm, which is essential for differentiating numerous tracks around the interaction point. Using these tracks, primary vertices are reconstructed with a resolution of up to 71 µm in beam direction and 13 µm transversal to it. This enables the impact parameter (IP)—the shortest distance between a track and a vertex—to be determined with a resolution of less than 35 µm for particles with a transverse momentum greater than 1 GeV/c.

The LHCb magnet is a dipole magnet [98] whose magnetic field is aligned to the y-axis, thus bending the tracks in the $x-z$ plane. The magnet field has an integrated magnetic flux of 4 T m and can be operated in two modes; the field is parallel to the y-axis in *MagUp* mode, while it is antiparallel in *MagDown* mode. During data taking, the polarity of the magnet is changed a few times per year. The SMOG datasets analysed in this thesis, however, are exclusively taken in MagDown mode. From the curvature of particle trajectories, momenta and charges can be determined.

The tracking system is completed by the TT and the three tracking stations T1-T3. Each station consists of four vertical layers, where the two innermost layers are tilted by a stereo angle of $\pm5^\circ$, providing precise information in both x and y. The TT is a silicon microstrip tracker that can reconstruct the trajectories low-momentum particles before they are deflected out of the detector acceptance by the magnet, as well as decay products of long-lived particles leaving not hits in the VELO. The tracking stations T1-T3 downstream of the magnet are composed of the Inner Tracker (IT) [99] and Outer Tracker (OT) [100, 101]. The IT has a cross-shape and includes the innermost region around the beam pipe, where particle density is expected to be the highest. Similar to the TT, the IT is a silicon microstrip tracker, which provides a spatial resolution of 48 to 55 µm [102]. It is covering an area of 120 cm × 40 cm, where the first number refers to width and the second number to height. The OT covers the remaining surface area of the tracking stations, occupying an area of 5 m × 6 m. The detector consists of three stations, each containing four double layers of straw tubes—amounting to twelve layers in total—providing a spatial resolution of 171 µm in Run 2 [103]. Charged particles passing through the straw tubes ionize the gas, thus creating signals that are recorded as hits, and drift time determination allows for precise spatial reconstruction.

Hits from various subdetectors are subsequently combined to reconstruct particle trajectories. Different types of tracks are identified, based on which subdetectors are used to reconstruct them. The different types of tracks are illustrated in Fig. 3.3. For the measurement presented in this thesis, only hits using the full detector information, called *long tracks*, are used, as they provide the highest precision; below a transverse momentum 20 GeV/c their momentum resolution is 0.5 %, going up to 1.0 % at 200 GeV/c.

Not every track corresponds to a real particle trajectory passing through the detector, as different hits in the detector material can be incorrectly combined. These incorrectly combined hits are known as *fake tracks* and can originate, for example, from high-occupancy events with many overlapping trajectories and in-flight decays of charged particles. To quantify the likelihood of a track being fake, the reconstruction software provides a neural-network-based variable [104], which is used to reduce the relative proportion of fakes. This variable will later be used in the presented measurement for the study of background contributions.

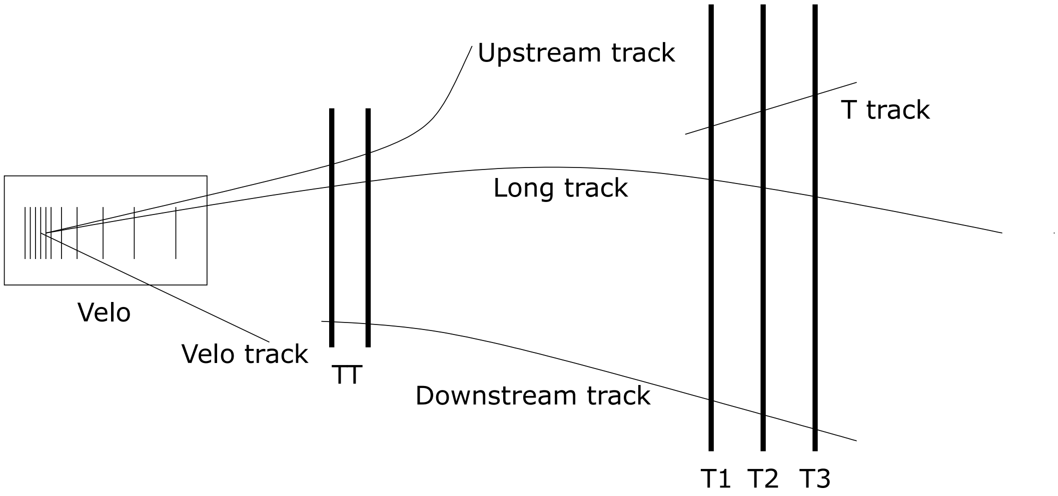


Figure 3.3 – Illustration of the LHCb tracking system and different track types, from Ref. [105]. Shown are the VELO and the four tracking stations; the different track types are illustrated via solid lines.

3.2.2 Particle identification

Particle Identification is made possible by the two RICH detectors, the calorimeter system, and the muon system. Using the RICH detectors [106], charged particles can be distinguished via the Cherenkov effect. Charged particles passing through a dielectric medium faster than the phase velocity of light in that medium—given by $c_\mathrm{n} = {}^{c}/_{n}$, with the medium specific refractive index n—emit electromagnetic radiation in a cone. The opening angle, θ, of the cone is a function of the refractive index and the particles' velocity, v, and given by $\cos\theta = \frac{c}{v\cdot n} = \frac{1}{\beta n}$. In the RICH detectors, these cones are reflected onto detection plates, resulting in ring images, from which the opening angles can be reconstructed. The ring images are associated to tracks, whose momenta are known. Based on this, likelihoods for different charged particles can be computed by applying mass hypotheses, thus enabling particle identification. However, the separation power of the Cherenkov-based PID in a specific medium is limited to specific momentum ranges. Due to this, the RICH1 and RICH2 use different mediums, allowing coverage of lower and higher momenta ranges. In the RICH1 detector, the fluorocarbon compound C_4F_{10} with a refractive index of $n = 1.0014$ is used, resulting in a momentum coverage of 2 to 60 GeV/c. In the RICH2 detector, CF_4 with refractive index of $n = 1.0005$ is used instead, covering the region 15 to 100 GeV/c. The reconstructed Cherenkov angle as a function of particle momentum, as measured by RICH1, is shown in Fig. 3.4. The RICH system shows strong separation power for the different charged particles, although certain momentum regions exist in which specific particles cannot be separated; below a momentum of 9.3 GeV/c, neither charged kaons nor protons emit Cherenkov light, thus making separation impossible. For each track, the reconstruction software provides delta log-likelihoods, DLLs, of particle hypotheses. For this, a pion mass is applied to each track and a pion log-likelihood is computed. Following this, the log-likelihoods for the proton, kaon, electron and muon mass hypotheses are computed and their difference to the charged pion hypothesis is provided.

The calorimeter system is used to determine the energy and impact position of hadrons, electrons and photons depositing their total kinetic energy. The main calorimeter system [107] consists of the ECAL and HCAL, which are based on the same design principle. Both are sampling calorimeters, consisting of alternating layers of scintillating and absorber material. In analogy to air showers, interactions with the absorber material initiate particle cascades, leading to the total deposition of their energy. The produced scintillating light is then measured by the scintillating layers, allowing the determination of the particle's energy. Although they share this general design, they are used for different processes. The ECAL is optimized for the identification of electron and photons. It uses lead as an absorber, providing a total thickness of 25 radiation lengths. The HCAL on the other hand is optimized for the identification of both charged and neutral hadrons. For this, iron is used as absorber material, resulting in a total depth of 5.6 nuclear interaction lengths. The main calorimeter system is supported by the SPD and PS detectors. The two subdetectors are separated by a thin lead layer, enabling additional separation power between electron- and photon-induced cascades; only electrons interact with the scintillating material in the SPD detector, while the lead plate causes photons to initiate an electromagnetic shower.

The PID system is completed by the muon system [108, 109], consisting of five muon stations, M1-M5. The first muon chamber, M1, is placed between the RICH2 detector the calorimeter system. It is a gas electron multiplier and only used to improve the transverse momentum measurement for the hardware trigger. The remaining muon chambers, M2-M5, are placed downstream the calorimeter system. They contain multiple multi-wire proportional chambers, dedicated to the measurement of charged particles that were not stopped by the calorimeter system. The only charged particles that should reach the M2-M5 are muons, due to their long life-time and minimal ionizing nature. However, due to the limited size of the calorimeter system, high-energetic hadrons could in principle reach the muon system. For this reason, the muon chambers are interlayered with 80 cm thick iron layers, reducing the misidentification probability to the percent level.

The only SM particles that can cannot directly be detected by the LHCb detector are neutrinos, as they do not interact with the detector material at all.

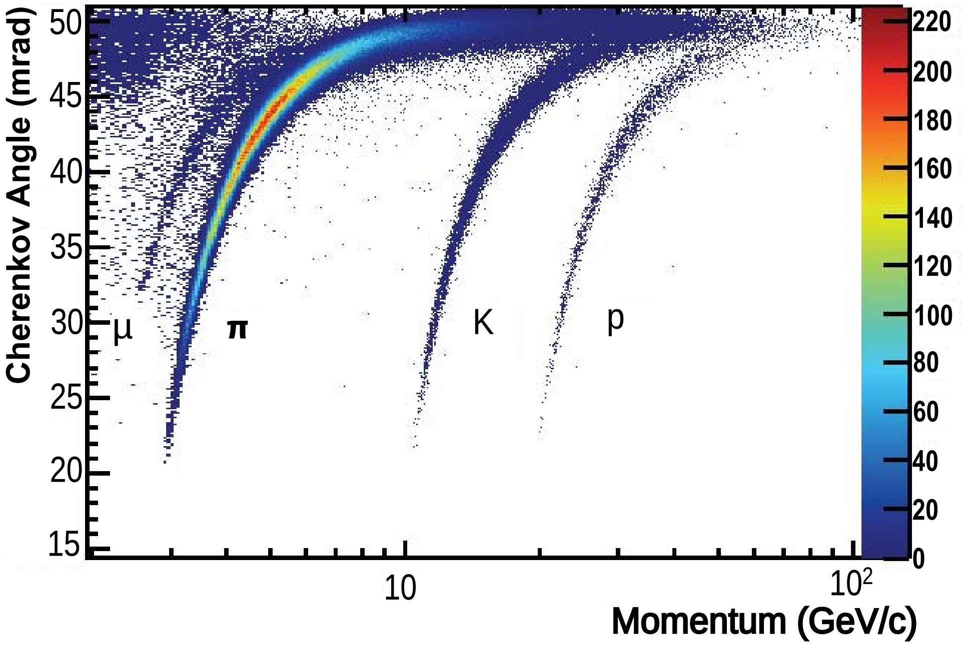


Figure 3.4 – Reconstructed Cherenkov angle for isolated tracks as a function of momentum, as measured by the RICH1 system, from Ref. [102].

3.2.3 System for the Measurements of Overlap with Gas

As previously mentioned, the LHCb experiment can also operate as a fixed-target experiment. This is possible due to the SMOG system [89]. The SMOG system was initially conceived for precise luminosity calibration [110] of colliding proton beams, for which it was installed in the LHCb detector in November 2011 during a technical stop [89]. Through the SMOG system, the noble gases helium, neon and argon can be injected into the vacuum vessel of the VELO. This injection creates a localized pressure bump extending approximately ±20 m around the interaction point, with a nominal pressure of about 2×10^{-7} mbar, which is two orders of magnitude higher than the nominal LHC vacuum pressure [95]. This was successfully exploited for precise luminosity determination using beam imaging techniques [111]. Between 2015 and 2018, the SMOG system was also used to study proton-nucleus and lead-nucleus collisions. For this purpose, several dedicated runs have already been performed with helium, neon and argon targets using proton and lead beams [95]. An overview of the collected SMOG datasets is given in Fig. 3.5.

A major benefit of this system are the easily interchangeable targets, allowing studies of diverse sets of proton-nucleon and proton-nucleon collisions. Due to its fixed-target configuration, this capability comes at the cost of a reduced nucleon-nucleon centre-of-mass energy, defined via

$$\sqrt{s_{\mathrm{NN}}} = \sqrt{2E_{\mathrm{N}} \cdot M_{\mathrm{N}}c^2 + M_{\mathrm{N}}^2c^4}, \tag{3.1}$$

where E_{N} is the beam energy and M_{N} the mass of the target per nucleus. While the latter is technically target-dependent, it is typically of the order of one proton mass. Therefore, a 6.5 TeV proton beam colliding with a fixed target corresponds to a nucleon-nucleon centre-of-mass energy of $\sqrt{s_{\mathrm{NN}}} = 110$ GeV.

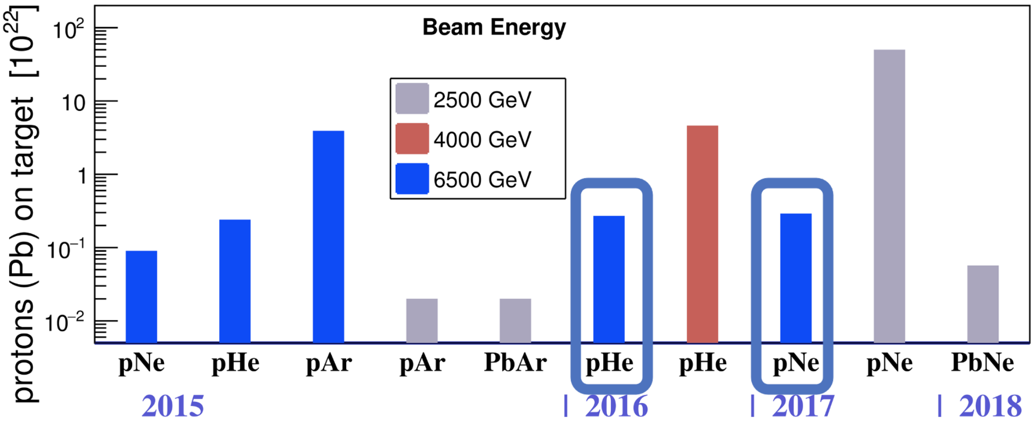


Figure 3.5 – Illustration of available SMOG datasets acquired during Run 2. The highlighted columns show the datasets used in this analysis. Picture modified from Ref. [95].

Within the detector upgrade for Run 3, the SMOG system was also upgraded to the SMOG2 [95] system. In this upgraded system, the gas is stored in a sophisticated storage cell, which is located upstream of the LHCb collision point. Outside the open ends of the cell, the gas density is suppressed by the VELO vacuum pumps, such that the beam-gas collisions occur mostly inside the cell. A sophisticated gas supply system allows a more accurate determination of the target pressure—and therefore luminosity—with the design pressure being two orders of magnitude higher than SMOG for a given gas [95, 112]. With this configuration, the SMOG system is no longer limited to noble gases, thus allowing the use of more gas species, such as hydrogen, deuterium, oxygen, nitrogen and the noble gases krypton and xenon [95]. Especially proton-oxygen and proton-nitrogen collision will be of great interest, as they are excellent proxies for air showers. Due to the increased gas pressure of the SMOG2 targets compared to SMOG, beam-gas interactions are more likely to occur, resulting in larger data samples. This makes SMOG2 measurements less prone to statistical uncertainties. First results are presented in Ref. [113] and several analyses are ongoing.

4 Measurement of cross-section ratios in proton-nucleon collisions

In this chapter, the measurement performed within this thesis is presented. It has not been published yet, but is currently undergoing internal review within the IFT working group of the LHCb experiment. In this measurement, double differential cross-section ratios of the strange hadrons K^0_S, $\mathit{\Lambda}$ and $\overline{\mathit{\Lambda}}$—collectively referred to as V^0, in analogy to their initial discovery in air shower experiments [7–10]—to charged pions, π^- and π^+, are performed. The measurement is performed utilizing the SMOG system, providing fixed-target proton-helium (p-He) and proton-neon (p-Ne) collisions recorded at a centre-of-mass energy of $\sqrt{s_\mathrm{NN}} = 110$ GeV. The measurement is initially performed in intervals of centre-of-mass rapidity, y^*, and transverse momentum, p_T, which is a standard two-dimensional phase-space grid used in differential cross-section measurements. For the study of strangeness enhancement, the measurement is additionally performed in intervals of transverse momentum and average multiplicity, $\langle N_\mathrm{ch}\rangle$, making it the first measurement of multiplicity-dependent strangeness production using the SMOG system.

4.1 Data and simulation samples

The data samples used in the presented measurement consist of p-He and p-Ne collisions. The samples were recorded with beam-energies of $E = 6.5$ TeV, resulting in centre-of-mass energies of $\sqrt{s_\mathrm{NN}} = 110$ GeV for each dataset. Both the p-He and p-Ne datasets were recorded during Run 2; the p-He dataset was recorded in 2016, while the p-Ne dataset was recorded in 2017. An overview of the collected SMOG datasets acquired during Run 2 is given in Fig. 3.5, in which the highlighted columns show the datasets used in this measurement. The p-He sample was collected over a period of four and a half hours and contains approximately 6×10^9 events. The p-Ne sample on the other hand was recorded over a period of 16 hours and contains approximately 1.7×10^{10} events. During data taking, both datasets were recorded using the same trigger configuration. On hardware level, all events are recorded at a fixed rate, independent of detector activity. At the first stage of the software level (HLT1), events are required to contain at least one reconstructed track within the VELO. At this stage, it is also checked that only events are recorded where a filled bunch of protons of the incoming beam intersects with an empty bunch from the outgoing proton beam. This type of event is referred to as *beam-empty* events.

For both datasets, a simulated sample containing approximately 10 000 000 events is provided. The initial generation of proton-nucleon events is done using `EPOS-LHC`, while the interactions of particles with the detector material and the detector responses are handled by the `GEANT4` toolkit [114, 115]. Using the same software used on data, tracks are reconstructed from the detector-response. The simulation allows linking of reconstructed tracks to generated particles; tracks that cannot be associated with generated particles are classified as fake. This information is later used to determine efficiencies and cross-checks. The detailed implementation will be discussed in the following chapters.

4.2 Analysis strategy

The goal of this analysis is to measure differential cross-section ratios of strange hadrons to charged pions, $K^0_\mathrm{S}/(\pi^+ + \pi^-)$ and $(\Lambda + \overline{\Lambda})/(\pi^+ + \pi^-)$. In doing so, only particles produced within a very narrow temporal and spatial distance from the primary interaction—where most of the energy is transferred—are taken into account. These particles are referred to as *prompt*. For the definition of prompt particles, this work follows the widely used definition of *primary* particles, as provided by the ALICE collaboration [116]. A particle is classified as primary if it is either directly produced in the interaction, or none of its ancestors are long-lived, meaning that no particle further up in the decay chain had a mean-lifetime exceeding 30 ps. An overview of long-lived particles is given in Tab. A.1. Following this definition, *primary* will be referred to as *prompt* in this thesis.

The double-differential cross-section with respect to two arbitrary variables, α and β, is defined as

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}\alpha\,\mathrm{d}\beta} \equiv \frac{n}{\mathcal{L}\cdot\Delta\alpha\cdot\Delta\beta}, \tag{4.1}$$

where $\mathcal{L}$ is the integrated luminosity, $\Delta\alpha$ and $\Delta\beta$ are the widths of the kinematic intervals in α and β, and n is the number of promptly produced particles observed within the intervals. The differential cross-section is additive, meaning that the yield of a composite category of particles is obtained by summing the yields of its components $n_\mathrm{A+B} = n_\mathrm{A} + n_\mathrm{B}$. In the above-mentioned datasets, the true count of particles is not directly accessible. Instead, candidates for V^0 particles and charged pions must be determined; V^0 candidates are reconstructed through their characteristic two-body decay topology, while long tracks are chosen as pion candidates. Details are given in Sec. 4.4 and Sec. 4.5. While the number of observed candidates is directly connected to the number of true promptly produced particles, several effects need to be taken into account. Due to the geometric acceptance of the LHCb detector and the finite efficiencies of the reconstruction algorithms, not every charged particle is reconstructed as a track, thus reducing the amount of observed candidates. For V^0 particles, an additional correction has to be applied to account for the branching fraction of the chosen decay channel. Furthermore, the measured particle count is also contaminated by background contributions that need to be subtracted. Consequently, the number of promptly produced particles can be written as

$$n = \frac{n_{\text{cand}} - \sum_i n_i^{\text{bkg}}}{\epsilon \cdot \mathcal{B}}. \tag{4.2}$$

Here, n_{cand} is the number of observed candidates, n_i^{bkg} is the contribution from background source i, ϵ is the total efficiency—given by the product of the geometric detector acceptance, the reconstruction, and selection efficiency—and $\mathcal{B}$ is the branching ratio. In case of a decay, the average values from the PDG are taken for the values of $\mathcal{B}$, however in the case of stable or directly detected particles, e.g. π^+ or π^-, the branching ratio is set to 1. The studied background sources for both V^0 hadrons and pions are non-prompt contributions, either from decays of long-lived particles or interactions with the detector material. The first background source that is specifically studied for pions are tracks that do not correspond to real particles, also known as fake tracks. The second pion-specific background source comes from particles that pass the particle identification (PID) requirements used to obtain a high-purity pion-sample. The first background source specific for V^0s is the combinatorial background, which refers to random combinations of tracks that did not originate from a common ancestor. The second specific V^0 source are cross-feed contributions from other V^0 particles, originate from wrongly assigned mass hypotheses during their reconstruction. A detailed description of how these contributions were accounted for is given in Sec. 4.4 and Sec. 4.5.

Inserting Eq. (4.2) into Eq. (4.1) results in the following expression for the double-differential cross-section:

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}\alpha\,\mathrm{d}\beta} = \frac{1}{\mathcal{L} \cdot \Delta\alpha \cdot \Delta\beta} \cdot \frac{n_{\text{cand}} - \sum_i n_i^{\text{bkg}}}{\epsilon \cdot \mathcal{B}}. \tag{4.3}$$

When taking the ratio of two such cross-sections, the luminosity and bin widths cancel, resulting in the following expression:

$$R(A, B)_{\alpha,\beta} = \frac{\mathrm{d}^2\sigma_A}{\mathrm{d}\alpha\,\mathrm{d}\beta} \Bigg/ \frac{\mathrm{d}^2\sigma_B}{\mathrm{d}\alpha\,\mathrm{d}\beta} = \frac{(n_{\text{cand, A}} - \sum_i n_i^{\text{bkg, A}}) \cdot (\epsilon_A \cdot \mathcal{B}_A)^{-1}}{(n_{\text{cand, B}} - \sum_i n_i^{\text{bkg, B}}) \cdot (\epsilon_B \cdot \mathcal{B}_B)^{-1}} \tag{4.4}$$

As most efficiency or background components are not known for real data, they have to be taken from simulation, which is reweighted to describe the data. However, even reweighted simulation does not describe the data with 100 % accuracy, which is why data-driven approaches are used to either determine efficiencies or correct efficiencies obtained from simulation.

The underlying analysis code is written entirely in the PYTHON programming language and a list of all the relevant software packages is given in Appendix A.2. In the following sections, the overall preprocessing of the samples are presented. This is followed by sections dedicated to the selection of V^0 particles and charged pions.

The same methods are applied to different collision and binning systems. To ensure clarity and avoid redundancies, a focus is set on intermediate results stemming from the p-He samples, in intervals of y^* and p_T; intermediate results from the p-Ne samples and $p_\mathrm{T} - \langle N_\mathrm{ch} \rangle$ intervals are not presented, unless significant differences are observed. Finally, the intermediate steps are combined, and the final results are presented.

4.3 Preprocessing and analysis concepts

In this chapter, steps of the analysis that concern both the V^0 and charged pion selection and observables, are discussed. This includes a global preselection after the software-trigger—the offline event selection—and application of the correct Lorentz-boost into the centre-of-mass system. In addition, the kinematic intervals used in the remaining analysis are defined.

4.3.1 Event selection

In order to study relative strangeness production, V^0 particles and charged pions need to be separated from background processes. The selection consists of several steps, namely the selection of the online trigger lines, the offline event selection, and the selection of the dedicated processes. The purpose of the event selection is to ensure that the event topology is well-defined, thus reducing ambiguities and removing events dominated by background processes. The event selection used in this thesis is inspired by the selection first introduced in Ref. [117]. It is applied to both data and simulation, independently of the studied particle of interest or studied collision systems. The applied event-selection is summarized in Tab. 4.1 and briefly explained in the following.

To ensure unambiguous association of tracks to a primary vertex, all events are required to have exactly one reconstructed primary vertex. The primary vertex must also fulfil two additional requirements. Firstly, the primary vertex is required to be reconstructed within the fiducial region, $\mathrm{PV}_z \in [-700, 100]\,\mathrm{mm}$, which is the established standard interval for SMOG analyses. In addition, the position of the primary vertex is also required to be consistent with the beam geometry. In order to reduce the contributions of spurious primary vertices from decays, secondary collision or combinatorial track matching, a cut is applied on the distance between the beam-centre and primary vertex at the same z-coordinate:

$$\sqrt{d_{xz}^2 + d_{yz}^2} < 400\,\mu\mathrm{m}, \tag{4.5}$$

with

$$d_{xz} = \mathrm{PV}_x - (\mathrm{offset}_{xz}/\mathrm{mm} - \mathrm{PV}_z \cdot \alpha_{xz}) \tag{4.6}$$

$$d_{yz} = \mathrm{PV}_y - (\mathrm{offset}_{yz}/\mathrm{mm} - \mathrm{PV}_z \cdot \alpha_{yz}), \tag{4.7}$$

Table 4.1 – Applied event selection on data and simulation.

Quantity	Requirement
Number of backward tracks	< 5
Number of primary vertices	$= 1$
z-coordinate of primary vertex	$\in [-700, 100]$ mm
Distance of primary vertex to beam centre at same z	$< 400\,\mu$m

where offset_{xz} and offset_{yz} are the beam-offsets and α_{xz} and α_{yz} are the inclination-angles of the beam profiles within the $x - z$ and $y - z$ planes, respectively. How these offsets and angles are determined in detail will be explored Sec. 4.3.2.

Finally, the number of backwards tracks N_{back}, i.e. tracks with negative pseudorapidity, is required to be smaller than 5. The cut is studied in detail in the internal documentation of Ref. [117] and is found to only remove a per-mille fraction of events while removing events with large multiplicities and many fake tracks, exceeding even the typical values for proton-proton collisions. This kind of events, called *splash events*, are believed to originate from beam-gas collisions up to 20 m upstream of the VELO. Particles from these collisions with high pseudorapidity may interact with the beam-pipe, thus creating high-energy showers. As a result of combinatorial track matching, *fake primary vertices* can be reconstructed; Similarly to fake tracks, these fake primary vertices are incorrectly reconstructed from random tracks which, by chance, mimic a real interaction point.

In addition to the general selection criteria described above, the IFT group has identified specific periods of data-taking with reduced quality. These periods, corresponding to subsets of the collected data, have been excluded from the measurement presented in this thesis. As a result, a total of 2.6×10^8 remain for the p-He data sample, while a total of 2.2×10^8 events remain for the p-Ne data sample.

4.3.2 Beam profiling

Apart from the separation from background, the correct treatment of the observables, as observed in the laboratory frame, is important. Due to the asymmetric collision, particles in the laboratory frame experience a mass-dependent boost along the z-axis. Therefore, to avoid species-dependent distortion in the ratios, the observables must be correctly Lorentz-boosted to the centre-of-mass system. However, simply Lorentz-boosting the observables along the z-axis assumes that the beam is perfectly aligned to it, which is not the case. Therefore, the correct approach is to account for the inclined beam by determining its inclination angles in the $x - z$ and $y - z$ planes. The four-momentum of a given particle can then be rotated according to Eq. (4.8), using the rotation matrices R_x and R_y to correct for the beam incline. Afterwards, the observables can safely be boosted into the centre-of-mass system.

$$\begin{pmatrix} E \\ p_x \\ p_y \\ p_z \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos -\alpha_{xz} & -\sin -\alpha_{xz} \\ 0 & 0 & \sin -\alpha_{xz} & \cos -\alpha_{xz} \end{pmatrix}}_{=R_x} \cdot \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos -\alpha_{yz} & 0 & \sin -\alpha_{yz} \\ 0 & 0 & 1 & 0 \\ 0 & \sin -\alpha_{yz} & 0 & \cos -\alpha_{yz} \end{pmatrix}}_{=R_y} \cdot \begin{pmatrix} E' \\ p'_x \\ p'_y \\ p'_z \end{pmatrix} \tag{4.8}$$

The beam is profiled using the mean positions of primary vertices in the $x - z$ and $y - z$ plane. The inclination angles and offsets are then determined by performing polynomial fits to the mean x and y values as a function of the PV_z positions. As the simulated beam conditions may not be identical to data, this is done separately for data and simulation. The fits in both planes are presented in Fig. 4.1 for p-He collisions, while the results for p-Ne are presented in Fig. 4.2. A linear fit is performed in the $y - z$ plane, while a parabolic fit is performed in the $x - z$ plane. For both fits, the constant term represents the vertical beam-offset, and the arctangent of the fitted slope corresponds to the inclination angle. The resulting offsets and inclination angles are summarized in Tab. 4.2 and Tab. 4.3.

Due to the small inclination angles, correcting for them has minimal impact on the observables and the difference between the proper treatment and a simple boost along the z-axis is much smaller than the uncertainties of this measurement. Nevertheless, this correction is carried out for the sake of consistency.

Table 4.2 – Obtained values for the beam inclination angles and offset after beam-profile fit for p-He data and simulation. For a quadratic model, two values are provided for the angle: the first for the lowest z-value, and the second for the highest z-value.

origin	plane	angle / rad	offset / mm
Data	$x - z$	$(-4.3698 \pm 0.0111) \times 10^{-4}$, $(-5.1070 \pm 0.0044) \times 10^{-4}$	$(8.1399 \pm 0.0005) \times 10^{-1}$
Simulation	$x - z$	$(-4.3051 \pm 0.0248) \times 10^{-4}$, $(-5.1037 \pm 0.0100) \times 10^{-4}$	$(8.0966 \pm 0.0011) \times 10^{-1}$
Data	$y - z$	$(-4.2266 \pm 0.0141) \times 10^{-5}$	$(-2.2537 \pm 0.0004) \times 10^{-1}$
Simulation	$y - z$	$(-4.2960 \pm 0.0347) \times 10^{-5}$	$(-2.1981 \pm 0.0010) \times 10^{-1}$

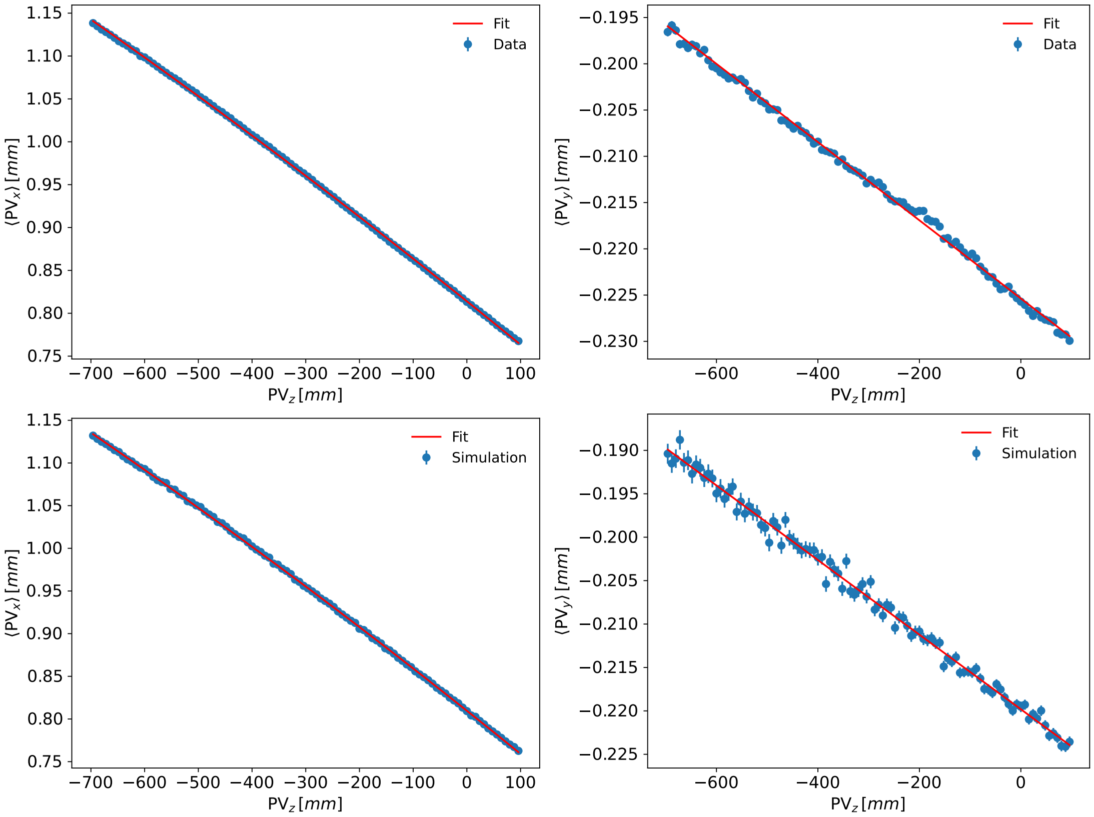


Figure 4.1 – Beam profiles for p-He collisions. The upper row shows the profiles in data, while the lower row shows the profiles in simulation. The left column shows the beam-profile in the $x - z$ plane and the right column shows the beam-profile in the $y - z$ plane.

Table 4.3 – Obtained values for the beam inclination angles and offset after beam-profile fit for p-Ne data and simulation. For a quadratic model, two values are provided for the angle: the first for the lowest z-value, and the second for the highest z-value.

origin	plane	angle / rad	offset / mm
Data	$x - z$	$(-5.1607 \pm 0.0117) \times 10^{-4}$, $(-4.8967 \pm 0.0047) \times 10^{-4}$	$(7.4440 \pm 0.0005) \times 10^{-1}$
Simulation	$x - z$	$(-5.0189 \pm 0.0143) \times 10^{-4}$, $(-5.0250 \pm 0.0057) \times 10^{-4}$	$(7.3753 \pm 0.0006) \times 10^{-1}$
Data	$y - z$	$(-6.0567 \pm 0.0141) \times 10^{-5}$	$(-2.0714 \pm 0.0004) \times 10^{-1}$
Simulation	$y - z$	$(-5.5828 \pm 0.0194) \times 10^{-5}$	$(-1.8709 \pm 0.0006) \times 10^{-1}$

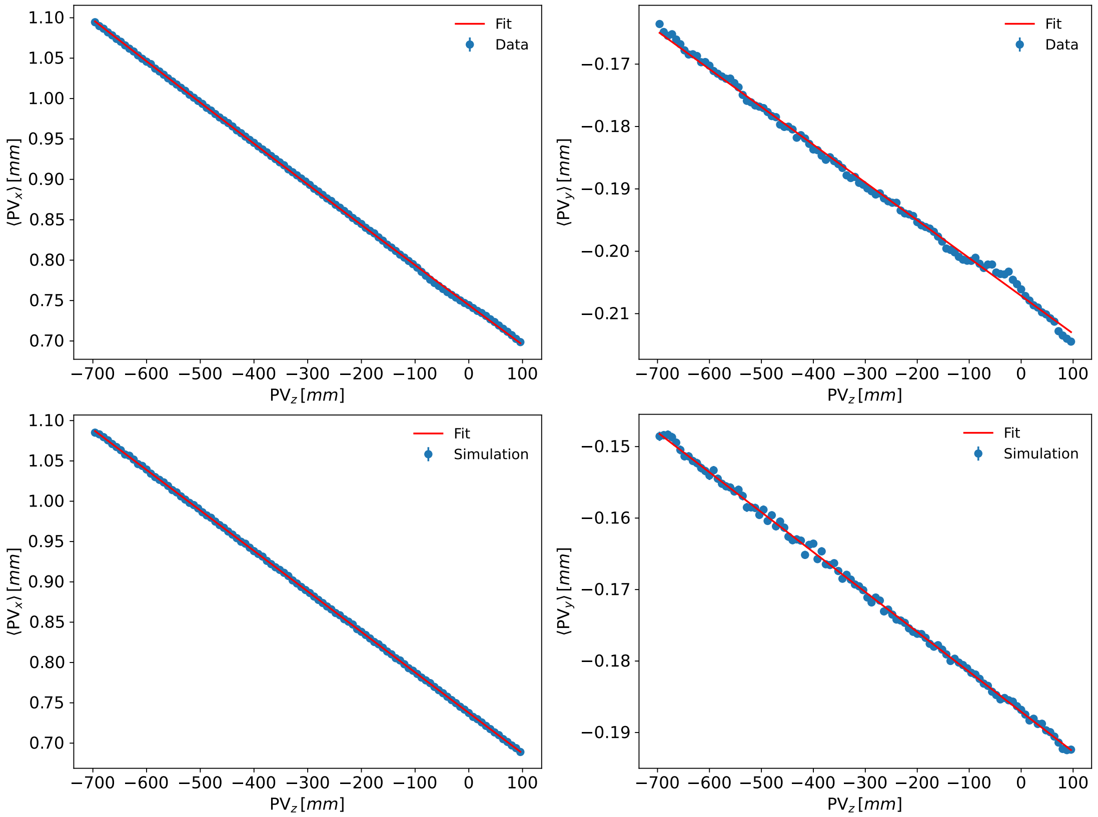


Figure 4.2 – Beam profiles for p-Ne collisions. The upper row shows the profiles in data, while the lower row shows the profiles in simulation. The left column shows the beam-profile in the $x - z$ plane and the right column shows the beam-profile in the $y - z$ plane.

4.3.3 Matching of simulated and real samples

For the accurate description of the detector-efficiency, the detector response needs to be simulated accurately. The LHCb `Tracking Group` found that the simulated tracking efficiency agrees well with data only if the detector occupancy is matched [118]. Consequently, the simulated samples must be reweighted to match the occupancy observed in data. This is especially important as the performance of tracking algorithms depends on the number of tracks and thus detector occupancy. The occupancy weight is determined by selecting an occupancy proxy and then comparing the number of events at each occupancy value in data and simulation. For this measurement, the number of long-tracks is chosen as occupancy proxy. The occupancy weight is then calculated as the ratio of the event count in data to that in simulation. The resulting weights for the *p*-He and *p*-Ne datasets are illustrated in Fig. 4.3. A clear difference is observed between the simulated *p*-He and *p*-Ne samples. The occupancy observed in the *p*-Ne samples is slightly higher than that in the *p*-He samples. Aside from this feature, the ratio for the *p*-Ne samples exhibits a flatter distribution than in the *p*-He sample, while the overall offset in both cases is proportional to the total number of events passing the selection in data.

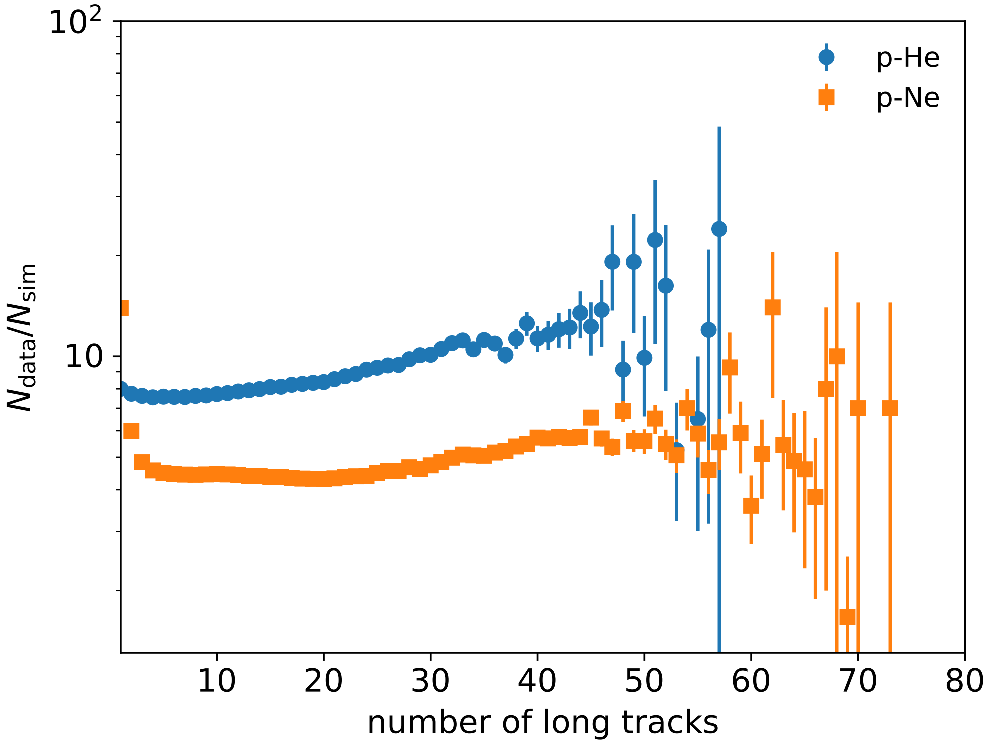


Figure 4.3 – Occupancy weights for the *p*-He and *p*-Ne samples.

4.3.4 Binning

As outlined in Chap. 4, the measurement is performed in two 2D systems, in kinematic intervals of $y^* - p_\mathrm{T}$ and $p_\mathrm{T} - \langle N_\mathrm{ch} \rangle$, for each of which the determination of the bins is explained in the following. The bin-edges chosen for the centre-of-mass rapidity are inspired by the known detector acceptance in pseudorapidity in the laboratory system, $\eta_\mathrm{lab} \in [2.0, 5.0)$. The rapidity range covered in the centre-of-mass system can be estimated using the relation

$$y' = \eta - y_\mathrm{cm}, \tag{4.9}$$

with

$$y_\mathrm{cm} \approx \operatorname{arsinh}\left(\sqrt{\frac{E_\mathrm{N}}{2 \cdot M_\mathrm{N} c^2}}\right), \tag{4.10}$$

where E_N is the beam energy and M_N the mass of the target per nucleus. For both the p-He and p-Ne datasets, y_cm is approximately 4.77. It is important to highlight that y' is not equal to the correctly Lorentz-boosted rapidity y^*, as it only gives a rough estimate based on the pseudorapidity, a geometric observable. Using this estimate, the interval in y^* is chosen to be $y^* \in [-2.5, 0.5)$. As the cross-section only varies slowly as a function of rapidity, six bins with a bin width of $\Delta y^* = 0.5$ are chosen.

Contrary to this, the differential cross-section in p_T follows a power-law distribution and covers multiple orders of magnitude. Therefore, a logarithmic binning is chosen instead, covering the interval $p_\mathrm{T} \in [10^2, 10^4)$ MeV/c. The number of bins—and hence the bin width—is chosen dependent on the studied 2D system. For the measurement in intervals of $y^* - p_\mathrm{T}$, 16 bins are chosen, corresponding to bin width of $\Delta \log(p_\mathrm{T}/\,\mathrm{GeV}/c) = 0.125$. To ensure sufficient statistics per bin, a broader binning of 4 bins is chosen for the measurement in $p_\mathrm{T} - \langle N_\mathrm{ch} \rangle$ intervals, corresponding to a bin width of $\Delta \log(p_\mathrm{T}/\,\mathrm{MeV}/c) = 0.5$.

Unlike linear or logarithmic binning, where bin edges are defined directly in the variable range, determining suitable intervals in average multiplicity requires a dedicated procedure that is applied separately for each dataset. Following the approach outlined by the ALICE collaboration in Ref. [31], five multiplicity classes are defined based on the relative occurrence of various amounts of long-tracks per event. Contrary to the definition employed by the ALICE collaboration, which uses ten classes, only five are employed here. This is done to compensate the reduced multiplicity ranges, and therefore narrow intervals, in fixed-target collisions compared to beam-beam interactions. Since ratios relative to charged pions, which make up the majority of long-tracks, are measured, simply counting all long-tracks could induce unwanted correlations between the observables and binning variable. For this purpose, two regions are defined. This is done using the estimated centre-of-mass rapidity y', as the mass composition of tracks is unknown in data. The measurement range, $y' \in [-1.0, 0.0)$, defines the region in which the analysis will be performed in, while the calibration range, $y' \in [-2.5, -1.0) \cup [0.0, 0.5)$, is used to determine the intervals in average multiplicity. The calibration range is chosen such that the track distribution is similar to the measurement range.

To determination of average multiplicity intervals involves several steps. First, the various amounts of tracks within the calibration range are counted. Using the percentiles listed in Tab. 4.4, five long-track intervals are determined. Using the reweighted simulated sample, the obtained intervals are mapped to the number of true prompt long-lived charged particles. As the simulation provides direct information on the species of true prompt long-lived charged particles, their true centre-of-mass rapidity can be used. Therefore, to implicitly compensate for the approximate Lorentz-boost, the number of tracks in the calibration range is mapped to the number of true prompt long-lived charged particles with centre-of-mass rapidity $y^* \in [-1.0, 0.0)$. The average multiplicity $\langle N_{\text{ch}} \rangle$ is then determined as the mean count of true prompt long-lived charged particles in the corresponding long-track interval, with negligible uncertainties. For both collision systems, the resulting mapping is illustrated in Fig. 4.4. This mapping is affected by tracking inefficiencies and the presence of non-prompt contributions. For example, a prompt V^0 particle can decay into two charged particles, resulting in two long tracks associated to non-prompt particles. In addition, due to inefficiencies in the event reconstruction, promptly produced charged particles are not guaranteed to be reconstructed as long-tracks, while the number of long-tracks can also include fake tracks. The resulting average multiplicities are additionally listed in Tab. 4.4.

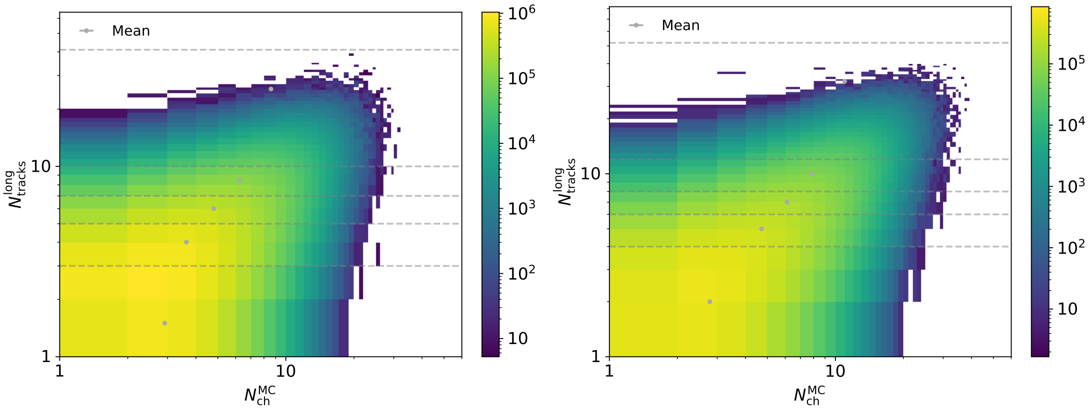


Figure 4.4 – Multiplicity mapping for the *p*-He (left) and *p*-Ne (right) datasets. The grey lines indicate boundaries of the long-track intervals, while the points represent the calculated, mean mapping of long-tracks to number of prompt charged particles.

Table 4.4 – Definition of the different multiplicity classes with their determined bin edges and average multiplicity per dataset.

class	$\sigma/\sigma_{\text{inel}}$	pHe		pNe	
		long-track intervals	$\langle N_{\text{ch}}\rangle$	long-track intervals	$\langle N_{\text{ch}}\rangle$
I	0 – 4.7 %	$[0,3)$	2.913 ± 0.001	$[0,4)$	2.782 ± 0.001
II	4.7 – 14 %	$[3,5)$	3.634 ± 0.001	$[4,6)$	4.714 ± 0.001
III	14 – 28 %	$[5,7)$	4.797 ± 0.001	$[6,8)$	6.106 ± 0.002
IV	28 – 48 %	$[7,10)$	6.249 ± 0.002	$[8,12)$	7.877 ± 0.002
V	48 – 100 %	$[10,41)$	8.594 ± 0.003	$[12,52)$	10.987 ± 0.004

4.4 Measurement of the prompt V^0 yield

In this section, the determination of the prompt V^0 yields is described. First, the reconstruction of V^0 candidates from specific decay channels is described. This is followed by a detailed description of the methods applied for background rejection and the estimation of residual background contamination. Subsequently, the estimation of the number of candidates in each kinematic interval as well as the determination of the total V^0 efficiency is presented. Systematic studies for the relevant steps are also outlined at the end of the respective sections.

4.4.1 Candidate identification and reconstruction

As V^0 particles are neutral, they do not produce hits within the detector. However, their average lifetime is sufficiently short that sufficiently many of them decay within the detector acceptance. The most likely decay a V^0 particle undergoes is into two charged hadrons; the K^0_S meson predominantly decays into two charged pions, π^+ and π^-, with a branching ratio of $(69.20 \pm 0.05)\,\%$ [36], while the Λ baryon decays predominantly into a proton and a negatively charged pion, with a branching ratio of $(64.1 \pm 0.5)\,\%$ [36]. Similarly to the Λ baryon, the $\overline{\Lambda}$ baryon decays into an antiproton and a positively charged pion with the same branching ratio [36]. The leading order Feynman diagrams, for the decays $K^0_S \to \pi^-\pi^+$ and $\Lambda \to p\pi^-$ are shown in Fig. 4.5. In this context, *leading order* refers to the simplest possible process, providing the dominant contribution.

As V^0 particles are not reconstructed by the used HLT1 trigger lines, there are no predefined objects containing the V^0 quantities in the used samples. Therefore, they must be manually reconstructed. This is done by combining track-pairs with opposite charge. In each event, all possible track pairs are determined and form a vertex candidate if oppositely charged. As the V^0 particle is expected to traverse some distance before decaying, a minimum distance along the z-axis of $\Delta z = 150\,\text{mm}$ between the primary vertex and the V^0 decay vertex is required in order to suppress combinatorial background. Following this, invariant mass hypotheses are calculated for each analysed V^0 particle. This is done by assigning mass hypotheses to final state

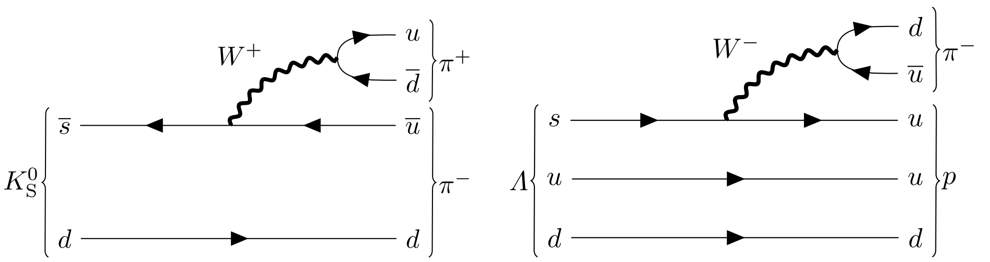


Figure 4.5 – Feynman diagrams of the decays $K^0_S \to \pi^-\pi^+$ (left) and $\Lambda \to p\pi^-$ (right) at leading order.

tracks according to the V^0 decay products. Once mass hypotheses are applied, the combined four momentum of the V^0 candidate can be calculated as the sum of the two tracks' four momenta, allowing the determination of the invariant mass. The invariant mass of a K^0_S candidate is required to be within 100 MeV/c^2 of the known K^0_S mass, and for Λ and $\overline{\Lambda}$ candidates, the invariant mass is required to be within 40 MeV/c^2 of their known mass. Once a candidate is selected, its four-momentum is corrected according to the procedure described in Sec. 4.3.2. The transformed four-momentum is then used in the subsequent analysis steps. The same procedure is applied to the final state candidates' tracks.

4.4.2 Candidate selection

When constructing V^0 candidates by combining opposite charged tracks, genuine signal processes are included; however, the likelihood of two unrelated tracks which happen to have opposite charges and fly close to one another is also high. This kind of background is known as *combinatorial background.* In addition, undesired contributions originating from either long-lived particles decaying into V^0 particles or interactions of particles with the detector material are also present and need to be accounted for. Finally, V^0 candidates may also contain cross-feed contributions from other V^0 particles. However, this contribution is discussed separately in Sec. 4.4.3.

For the reduction of background processes, six abstract, geometric variables are employed. The derivation of these variables is presented internally within the LHCb collaboration by Michael Schmelling in Ref. [119] and is summarized in Appendix A.3. The variables can be organized into two groups of three variables. The first group contains the distance of closest approach between the V^0 decay products, D, the distance from the primary vertex to the event plane, S, and the in-plane impact-parameter of the V^0 candidate to the primary vertex, I. For promptly produced V^0 particles, these three quantities are expected to be close to zero. The second group includes the in-plane squared impact parameters of the tracks from the primary vertex, I_{11} and I_{22}, and a variable containing the information on the flight-distance not covered by the other variables, I_{12}, which are all sensitive to non-zero lifetimes.

The distributions of the six geometric variables for $K^0_S \to \pi^-\pi^+$ candidates in p-He collisions are illustrated in Fig. 4.6. The distributions are shown for prompt K^0_S particles,

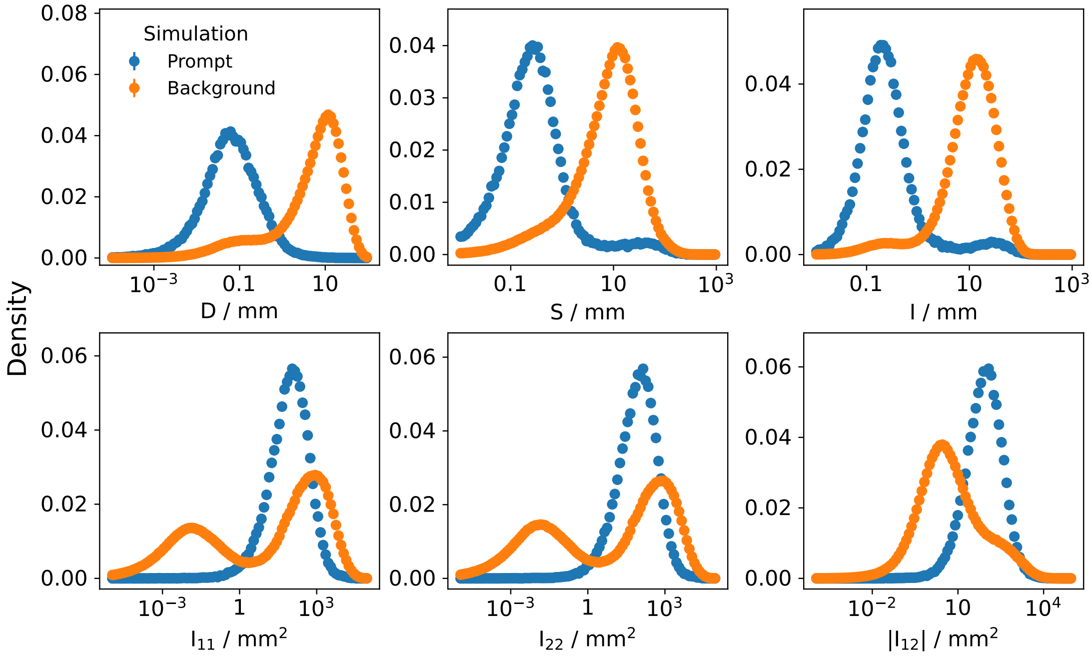


Figure 4.6 – Distribution of BDT input variables for $K^0_S \to \pi^+\pi^-$ candidates in the p-He simulated sample. Here, signal refers to prompt K^0_S particles, whereas background refers to everything else.

which are labelled as *Signal*, while remaining contributions from either non-prompt K^0_S particles, combinatorial and cross-feed contributions are labelled as *Background*. The D, S and I distributions show a clear separation between signal and background distributions. This is expected since these variables are close to zero for the two-body decays of promptly produced V^0 particles, while they tend to be larger for combinatorial backgrounds and non-prompt components. The variables I_{11}, I_{22} and I_{21} show a peaking structure for promptly produced V^0 particles; for I_{11}, I_{22}, the background contribution show a double-peak structure. Using the simulated sample it has been verified that both peaks arise from combinatorial background, with the second peak receiving an additional contribution from non-prompt K^0_S particles. The distributions for Λ and $\overline{\Lambda}$ candidates are presented in Appendix A.4. Overall, the trends for Λ and $\overline{\Lambda}$ baryons are similar, but offer less separation power. This has been checked using the simulated sample and is mainly caused by an increased non-prompt contribution, caused by decays of long-lived strange hadrons. The distributions of I_{11}, I_{22} are interchanged for Λ and $\overline{\Lambda}$ baryons. This effect arises from the internal ordering of tracks, which directly impacts the variables; the first track is always negatively charged, while the second track is always positively charged. No major differences are observed between the p-He and p-Ne datasets, which is expected due to the purely geometric nature of these variables.

The six geometric variables are a linearly independent set for two-body decays and are therefore perfectly suited as input for a machine learning classifier. Machine-learning classifiers typically perform better when input variables are on a similar numerical scale. However, the six variables span several orders of magnitude. Therefore, the logarithm of these variables is taken as input instead. This is not an issue, as all variables except I_{21} are strictly positive, for this latter problematic case the logarithm of the absolute value is taken instead.

For this measurement, a boosted decision tree (BDT) is employed as machine-learning classifier. It is implemented using the XGBClassifier provided the XGBoost library [120]. To mitigate the effect of class imbalance between signal and background in the training dataset, the *scale_pos_weight* parameter of the XGBClassifier is used. This scales the contributions of the signal in the loss-function according to the ratio

$$r = \frac{N_{\text{Background}}}{N_{\text{Signal}}}, \tag{4.11}$$

with the number of background entries, $N_{\text{Background}}$, and the number of signal entries, N_{Signal}. The maximum depth of the BDT is chosen as 4, while the *logloss* function is chosen for the evaluation metric, as it is a common choice for binary classification. The classifier response can be any value between 0 and 1, with 0 being most likely background and 1 being most likely signal. It can therefore be interpreted as a probability of a V^0 candidate being a true promptly produced V^0 particle.

A common approach when employing a BDT is to divide the available simulated sample into a training and testing subset. However, this renders the training subsample unusable in the subsequent analysis steps, as applying the BDT on data it has been trained on could introduce a bias. To avoid reducing the amount of simulation available for later analysis steps, *k*-folding is employed. The basic principle of k-folding is to split the trainings sample into k equal subsets or *folds*. This way, k classifiers can be trained on $k-1$ subsets each, with the remaining subset being available for evaluation. By doing so, the majority of data can be used for the training of the classifiers. As each classifier can safely be applied to the sub-set it was not trained on, the complete simulated dataset remains available for subsequent measurement steps. The application of the classifiers on data is then slightly different. Since all classifiers are equally valid and have not been trained on data before, the mean of all classifier responses is used instead. In this analysis, *k*-folding is employed with $k = 5$.

The classifier responses for each analysed V^0 decay are illustrated in Fig. 4.7, for one sub-set of the simulated p-He sample each. The classifiers show a clear separation of signal and background for each V^0 particle; the background is always peaking near 0, while the signal is always peaking near 1. Despite the number of background entries being significantly higher than the number of signal entries, the classifier performance remains high and robust, indicating strong discriminating power for each V^0 particle. For K^0_S candidates, the classifier assigns a roughly equal amount of signal and background to a score of 0.4. Using the simulation truth information, it is observed that the background excess in this region is caused by the non-prompt component, which for K^0_S particles consists exclusively of material interactions. For both $\mathit{\Lambda}$ and $\overline{\mathit{\Lambda}}$ candidates, the number of signal entries with a low classifier response differs between the trainings and test datasets. Judging by the error bars and general shape of the curves, the limiting factor seems to be the total number of Λ and $\overline{\Lambda}$ particles available in the simulated sample. This is not a problem, as these regions will be excluded by applying a cut on the classifier response. This will be discussed in more detail further below. For each V^0 particle, the number of background entries starts to peak again for values near 1. Truth-level information from the simulation indicates that the background in this region is dominated by the non-prompt component, with additional contributions from the cross-feed and combinatorial components. Generally, the number of background entries with a classifier-response close to 1 relative to the signal entries is small, and therefore not a problem. However, the relative non-prompt contribution is visibly higher for Λ and $\overline{\Lambda}$ candidates than for K^0_S candidates and therefore requires additional attention, which will be addressed at the end of this section.

The strong separation power is also indicated by the receiver operating characteristic (ROC) curves, which are, for one subset of the simulated p-He dataset, illustrated in Fig. 4.8 for each analysed V^0 particle. A ROC curve shows the false-positive rate on the x-axis versus the true-positive rate on the y-axis. The false-positive rate, FPR, is the fraction of background that is misidentified as signal and is defined as

$$\mathrm{FPR} = \frac{\#\mathrm{False\ \ Positive}}{\#\mathrm{False\ \ Positive} + \#\mathrm{True\ \ Negative}}, \tag{4.12}$$

while the true-positive rate, TPR, is the signal efficiency and defined as

$$\mathrm{TPR} = \frac{\#\mathrm{True\ \ Positive}}{\#\mathrm{True\ \ Positive} + \#\mathrm{False\ \ Negative}}. \tag{4.13}$$

A *True Positive* occurs when a signal entry is correctly identified as such, while a *False Positive* refers to a background entry that is, incorrectly, identified as signal. Similarly, *True Negative* refers to a background entry that is correctly identified as such, while *False Negative* refers to a signal entry that is, incorrectly, classified as background. Therefore, strong separation power is indicated by a signal efficiency close to 1 while having a false-positive rate close to zero. This information can be summarized as a single number, the Area under Curve (AUC), which is 1 for perfect separation, 0.5 for no separation or random guessing, and 0 for total inversion, meaning a strong separation of signal and background which are simply labelled incorrectly. As the AUC is close to one for each V^0 species, the classifiers seem to be able to separate prompt V^0s from background sources.

For the remainder of the analysis, a cut on the classifier response is applied for each V^0 species. The cut is chosen such that the majority of signal is kept, while suppressing as much background as possible; the cut values are determined such that the signal efficiency is 0.9.

As mentioned earlier, the increased relative non-prompt components in Λ and $\overline{\Lambda}$ candidates requires additional treatment. Studies using the simulated sample revealed that a significant contribution stems from long-lived, multi-strange particles. To suppress this contribution, an additional constraint is put on Λ and $\overline{\Lambda}$ candidates by requiring the impact parameter to be smaller than 0.12 mm. This variable is chosen as it generally provides strong separation power between no prompt and non-prompt component. However, in case of Λ and $\overline{\Lambda}$ baryons, no clear separation can be achieved. Therefore, this cut has been manually optimized to increase the purity of the remaining sample while preserving sufficient statistics.

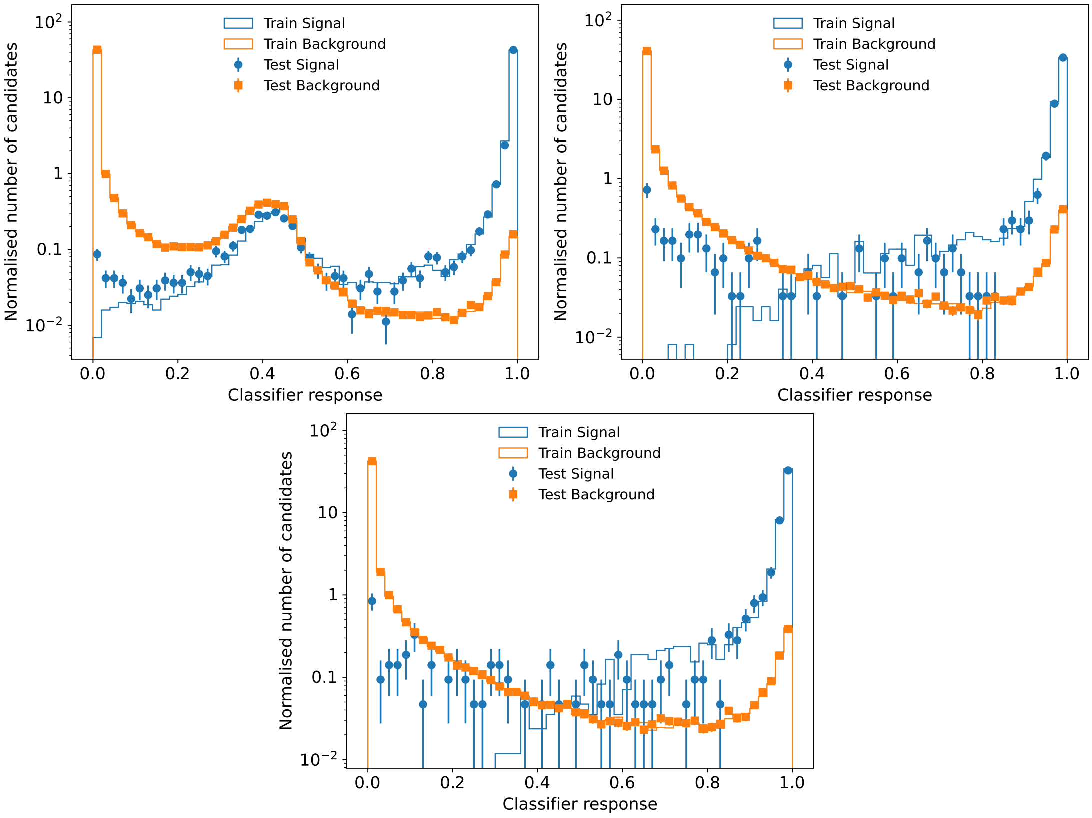


Figure 4.7 – Classifier response for K^0_S (upper left), Λ (upper right), and $\overline{\Lambda}$ (lower middle) candidates in the first of five subsets of the simulated *p*-He sample. The label *Signal* refers to the promptly produced particle of interest, while *Background* refers to everything else.

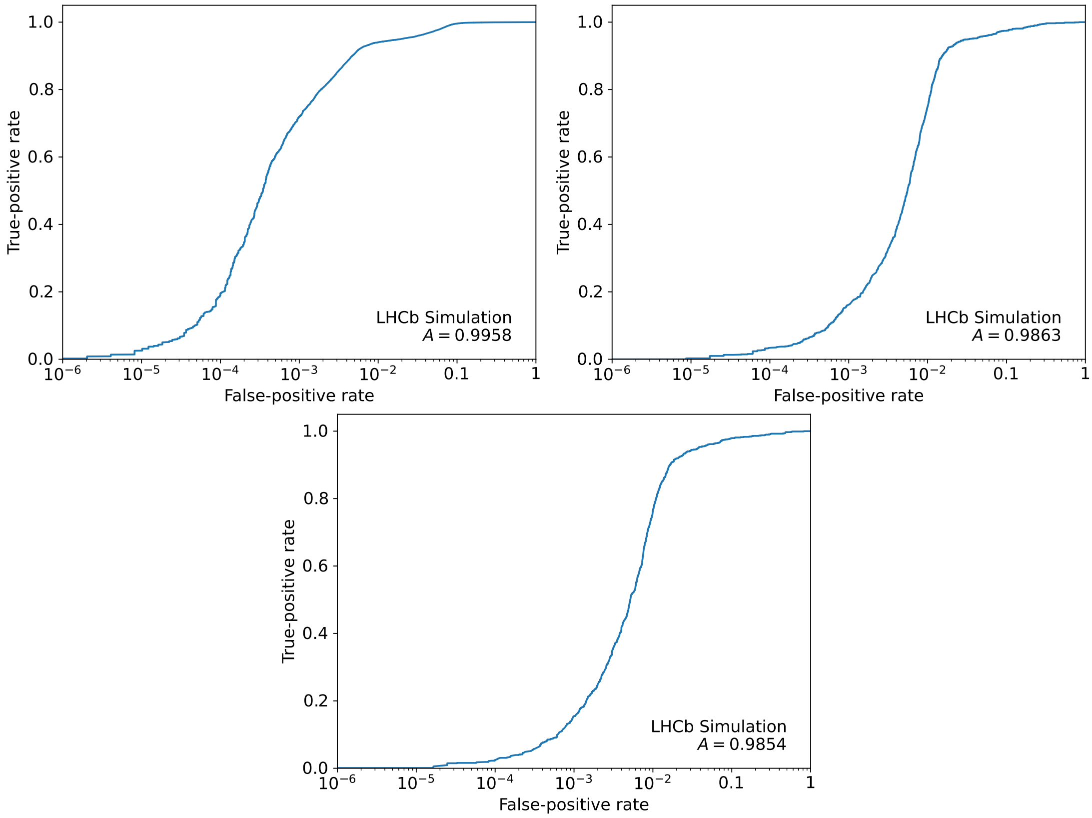


Figure 4.8 – ROC curve for K^0_S (upper left), $\mathit{\Lambda}$ (upper right), and $\overline{\mathit{\Lambda}}$ (lower middle) candidates in the first of five subsets of the simulated p-He sample.

4.4.3 Mass veto

The BDT selection and the requirements on the impact parameter applied in the previous section significantly suppress combinatorial and non-prompt contributions. However, they do not inherently distinguish between $\Lambda \to p\pi^- + cc.$ and $K^0_S \to \pi^+\pi^-$ decays. The reconstruction of these two-body decays differs only in the assigned mass-hypothesis of the final-state tracks. Therefore, assigning an incorrect mass hypothesis to one of the final-state tracks can cause a decay of one particle species to populate the reconstructed mass region of another. These cross-feed contributions are not suppressed by the applied BDT, as its input variables are purely geometric in nature. As a result, these cross-feed contributions must be explicitly vetoed.

For K^0_S particles, the veto is performed by assigning the proton mass-hypothesis to either of the final state tracks. If the resulting invariant mass lies within 9 MeV/c^2 of the known Λ mass, the candidate is vetoed. The same procedure is applied for Λ and $\overline{\Lambda}$ candidate by assigning a pion mass hypothesis to the corresponding proton track candidate. If the resulting invariant mass lies within 20 MeV/c^2 of the known K^0_S mass, the candidate is vetoed. The recomputed invariant masses of K^0_S mesons and Λ baryons from the p-He simulated sample are visualized in Fig. 4.9. There, the cross-feed of true Λ and K^0_S mesons are explicitly highlighted. The remaining cross-feed contribution after applying the cut is negligible and thus not studied further.

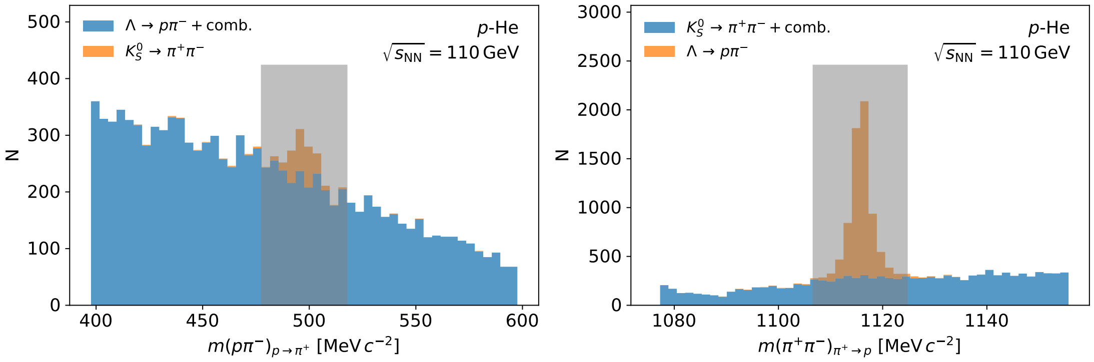


Figure 4.9 – Invariant mass of V^0 candidates with substituted mass hypotheses in the p-He simulated sample. The substituted mass hypothesis $p \to \pi^+$ in the Λ sample is illustrated in the left plot, while the $\pi^+ \to p$ substitution in the K^0_S sample is illustrated in the right plot. The cross-feed of true K^0_S and Λ particles is highlighted in orange. The blue distributions show the remaining invariant mass of the respective sample, i.e. true Λ (left) and K^0_S (right) particles and combinatorial background. The grey areas highlight the vetoed regions.

4.4.4 Signal yield extraction via invariant mass fits

After applying the selection and veto outlined in the preceding two sections, the next step is to determine the number of V^0 candidates. This is done by performing binned extended maximum-likelihood fits [121] to both data and simulation. The primary goal of the fits is to separate the V^0 candidates from combinatorial background. As prompt and non-prompt V^0 particles are indistinguishable at this stage, they are collectively treated as signal. The fits are performed in every kinematic interval with sufficient statistics, requiring 100 candidates in both data and reweighted simulation.

The V^0 signal component is fitted using a non-standardized Student's t distribution [122], which is an interpolation between a non-relativistic Breit-Wigner and a Gaussian distribution. A non-standardized Student's t distribution has the following three parameters: the mean value, μ, the standard deviation s, and the number of degrees of freedom, ν, which serves as shape parameter. The latter is originally defined as a discrete variable, but is allowed to vary continuously during the fit. To increase numerical stability, the value $1/\nu$ is fitted instead of ν. The student's t function is normalized to 1 and then scaled by the signal count N_{V^0}.

Given the absence of any background source other than the combinatorial background, the background component is modelled using a second-order Bernstein polynomial. While these polynomials possess the excellent numerical properties of Chebyshev polynomials, they come with the advantage that their density cannot be negative as long as all parameters are non-negative. This avoids numerical issues during the minimization of the cost-function.

To mitigate background-induced bias, only the combinatorial background is fitted first. This is done by excluding all entries within a narrow window around the expected signal peak. The resulting background fit parameters are then fixed, and the signal peak is fitted separately within the narrow window. In the subsequent steps, the total model, i.e. signal plus background, is determined. To ensure convergence when the signal peak is not clearly defined, weak Gaussian constraints are employed in addition to parameter constraints. The obtained complete model is then compared to the background-only model. The total model is only chosen when a fit with a non-zero signal peak performs significantly better than the background-only model. This improvement is quantified by the difference in the log-likelihoods, which is required to be at least 10.

This procedure is then used to obtain the statistically independent K^0_S, $\mathit{\Lambda}$ and $\overline{\mathit{\Lambda}}$ counts in every kinematic interval. Representative fits of the invariant mass distribution for K^0_S, $\mathit{\Lambda}$ and $\overline{\mathit{\Lambda}}$ candidates in both data and simulation are illustrated in Fig. 4.10 for chosen $y^* - p_T$ intervals.

To check whether the performed fits are in any way biased, they are also performed on the occupancy-reweighted simulated sample. For every kinematic interval, the fitted V^0 count is compared to the true number of reconstructed V^0 particles. The absolute difference between these two values is then used as relative systematic uncertainty. The K^0_S, $\mathit{\Lambda}$ and $\overline{\mathit{\Lambda}}$ counts obtained for the real p-He sample, including their statistical and systematic uncertainties are illustrated in Fig. 4.11, Fig. 4.12 and Fig. 4.13, in intervals of $y^* - p_T$. For K^0_S particles, the statistical uncertainty is below 3 % for the majority of intervals, increasing to 20 % towards the kinematic edges. The systematic uncertainty is below 4 % for the majority of intervals. Large deviations of up to 20 % are mostly observed at the lower or upper edges, and thus originate from low statistics. For $\mathit{\Lambda}$ and $\overline{\mathit{\Lambda}}$ particles, the statistical uncertainties are slightly increased, being between 1.5 and 10 % for the majority of intervals. Towards the kinematic edges, the uncertainties increase

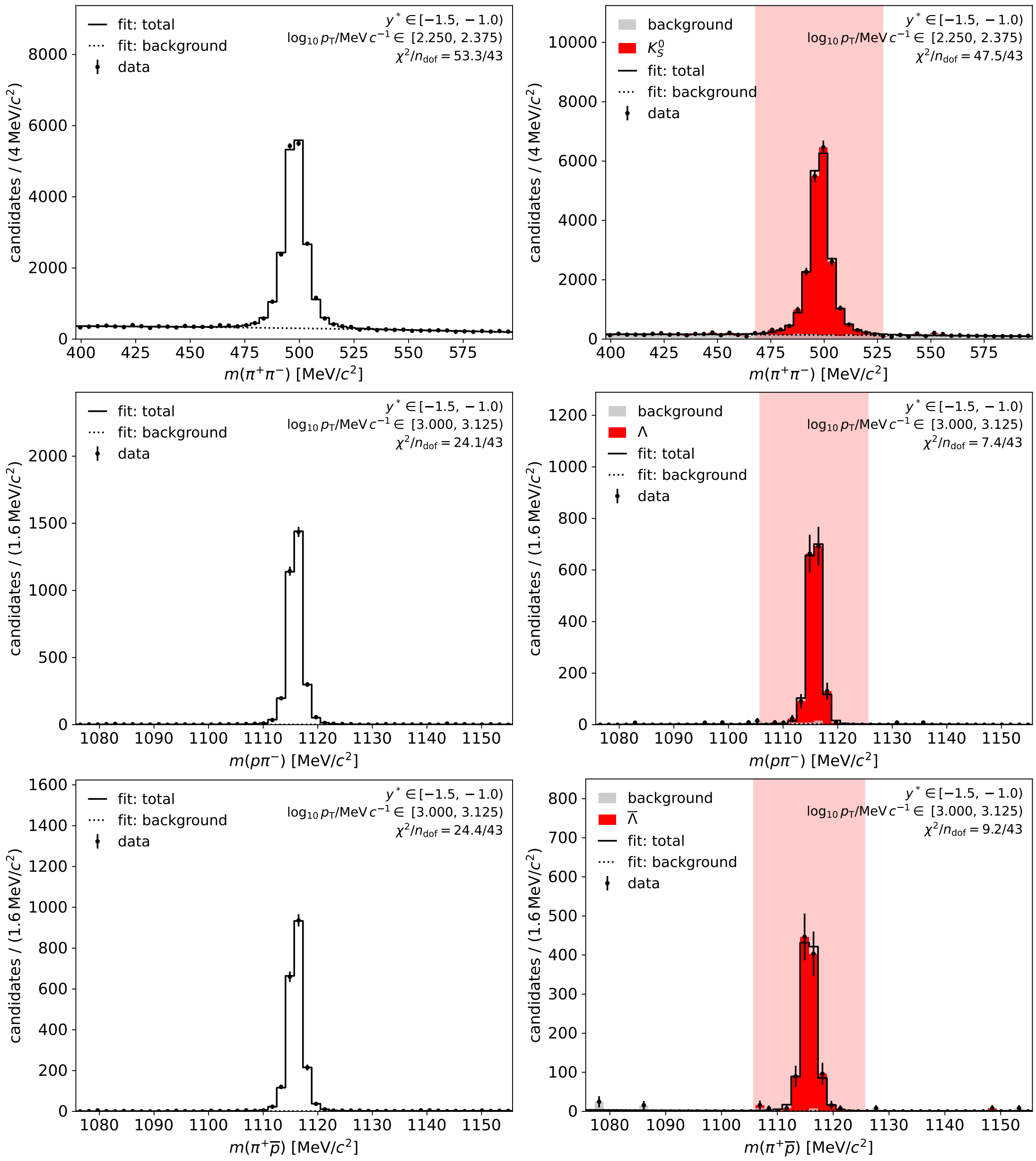


Figure 4.10 – Fitted invariant mass distributions obtained from the real (left) and simulated (right) p-He samples for chosen $y^* - p_T$ intervals. The distributions are presented for K_S^0 (top row), Λ middle row and $\overline{\Lambda}$ (bottom row) candidates. The solid line shows the fitted complete model while the dashed line shows the background-component of the model. For the simulated samples, the background and particle distributions are tracked separately and stacked for illustration.

up to 30 %. In addition, the number of intervals in which a fit could be performed is reduced compared to K^0_S particles. The systematic uncertainties are below 7 % for the majority of intervals, increasing towards the kinematic edges, where individual intervals exceed 50 %. Again, this is likely caused by the reduced statistics in these intervals.

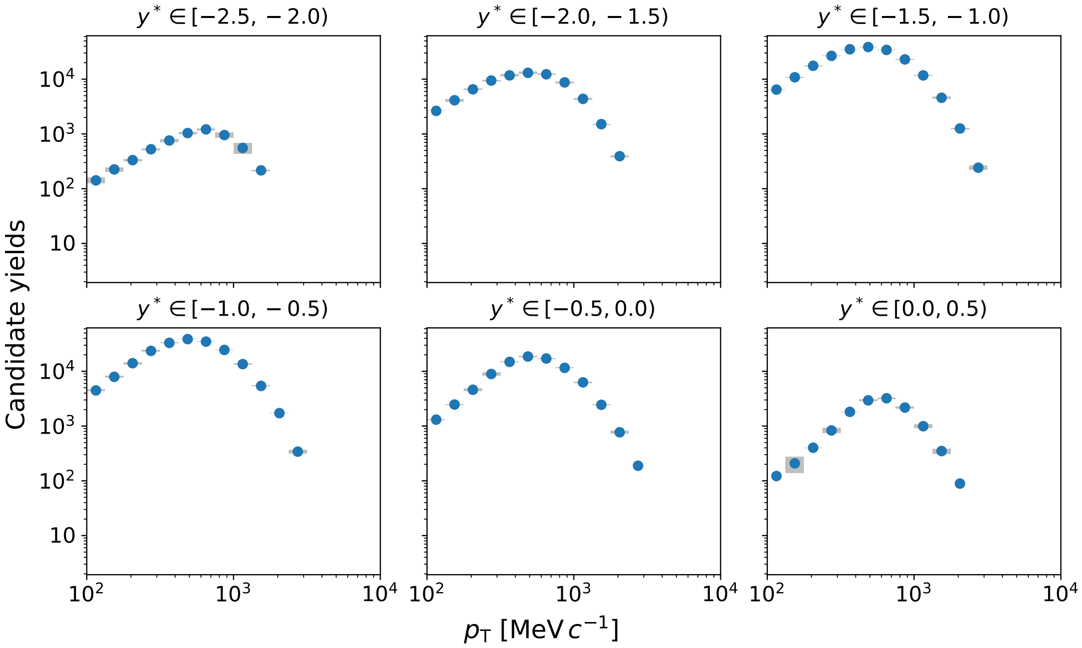


Figure 4.11 – Obtained K^0_S counts for the real p-He sample in intervals of $y^* - p_\mathrm{T}$. The bars and boxes represent the statistical and systematic uncertainty, respectively.

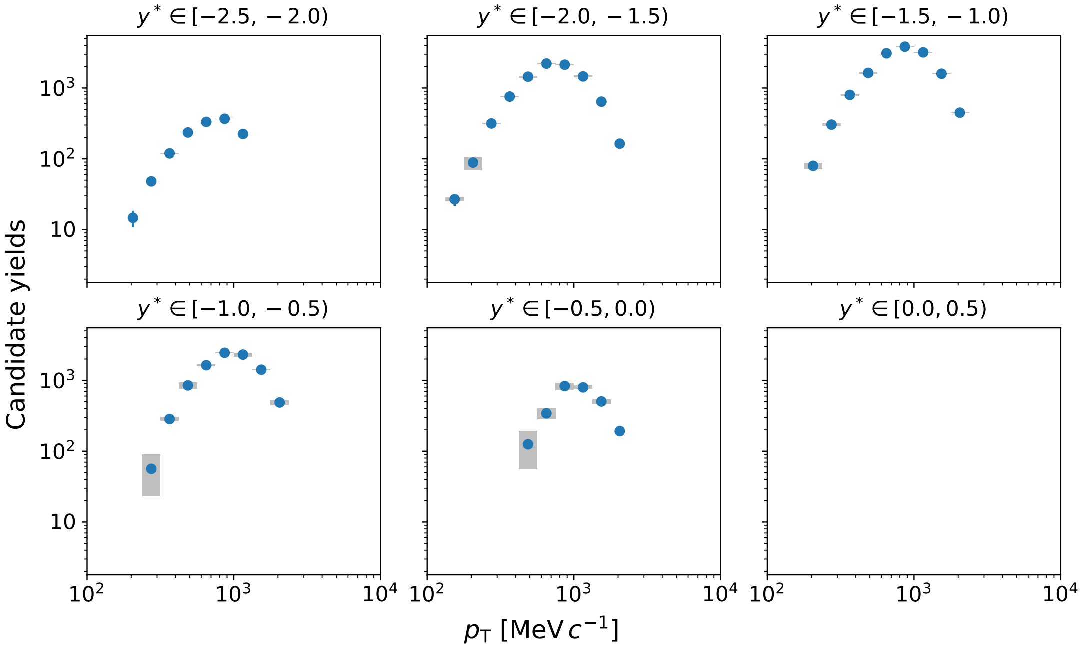


Figure 4.12 – Obtained Λ counts for the real p-He sample in intervals of $y^* - p_T$. The bars and boxes represent the statistical and systematic uncertainty, respectively.

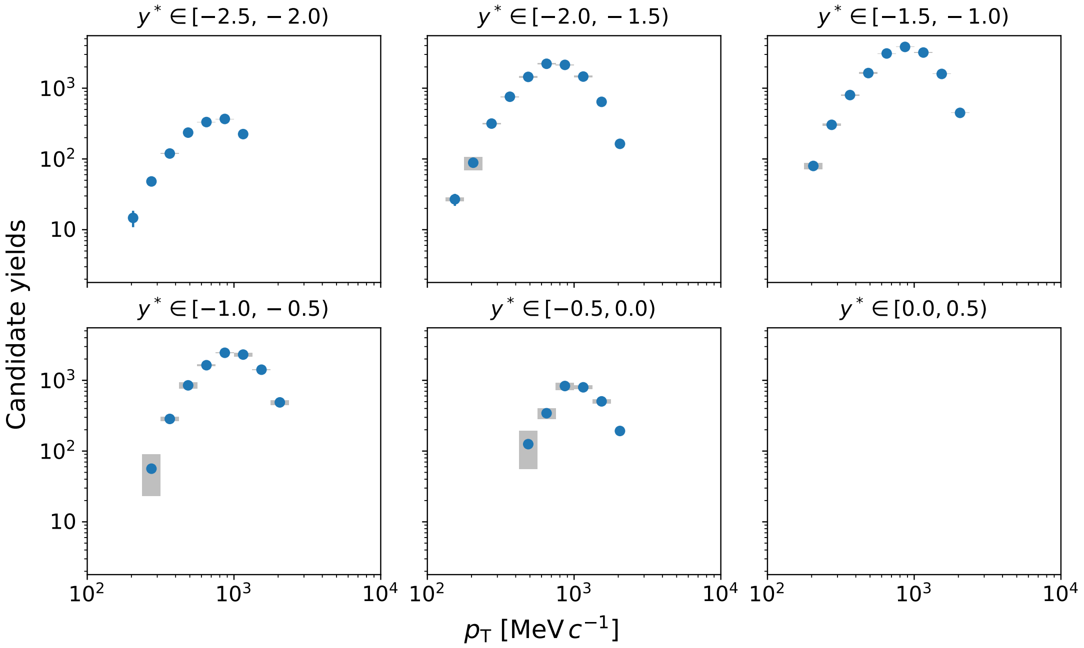


Figure 4.13 – Obtained $\overline{\Lambda}$ counts for the real p-He sample in intervals of $y^* - p_T$. The bars and boxes represent the statistical and systematic uncertainty, respectively.

4.4.5 Remaining background contributions

As the fits of the invariant-mass distribution discussed in Sec. 4.4.4 already successful separate K^0_S, Λ and $\overline{\Lambda}$ particles from combinatorial background, only the remaining contamination from the non-prompt components needs to be determined. As outlined earlier, this contamination can either stem from decays of long-lived particles or from interactions with the detector material.

The origins of true K^0_S, Λ and $\overline{\Lambda}$ particles in the simulated p-He sample are presented in intervals of $y^* - p_T$ in Fig. 4.14, Fig. 4.15 and Fig. 4.15, respectively. As illustrated in Fig. 4.14, non-prompt K^0_S particles only stem from material interactions. However, their relative contribution is small for the majority of intervals. The compositions of Λ and $\overline{\Lambda}$ presented in Fig. 4.15 and Fig. 4.16 show a similarly small contamination from material interactions. However, there is an additional contribution from strange hadron decays, which is already suppressed by the cut on the impact parameters applied in Sec. 4.4.2. The kinematic intervals with contributions exceeding 10% are located at either the lower or upper edges and are therefore constrained by statistics, making them less of a concern. Since the background contributions for the remaining intervals are minimal, they will not be investigated further. Their impact is instead accounted for by assigning a 50 % systematic uncertainty to the background contribution derived from the simulated sample.

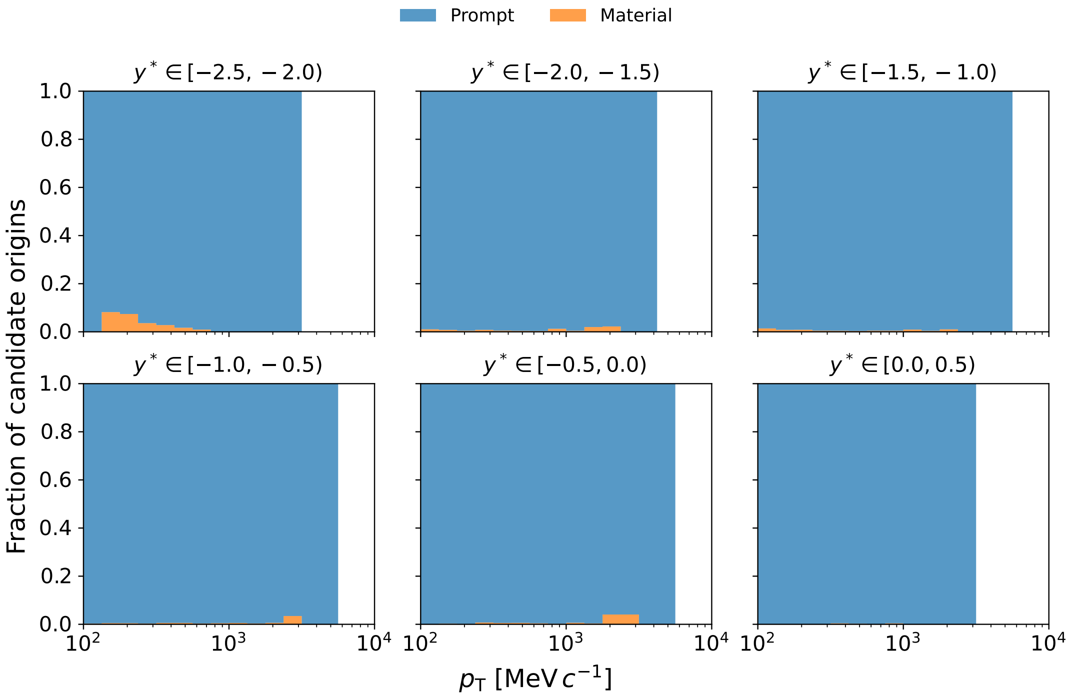


Figure 4.14 – Origins of K^0_S particles in the simulated p-He sample.

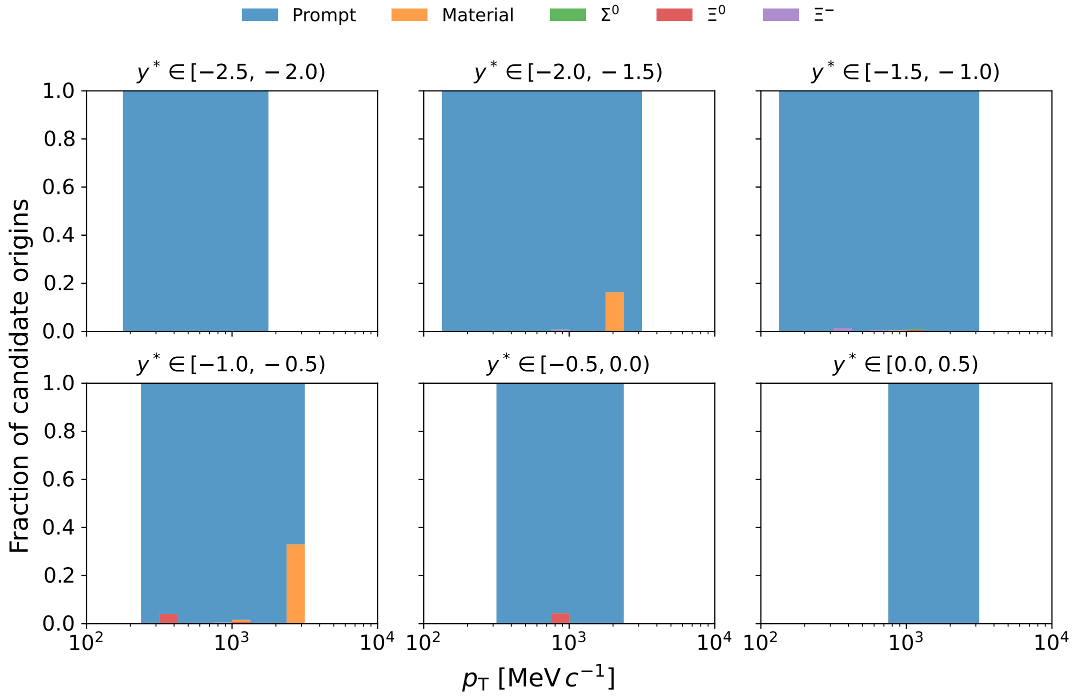


Figure 4.15 – Origins of Λ particles in the simulated *p*-He sample.

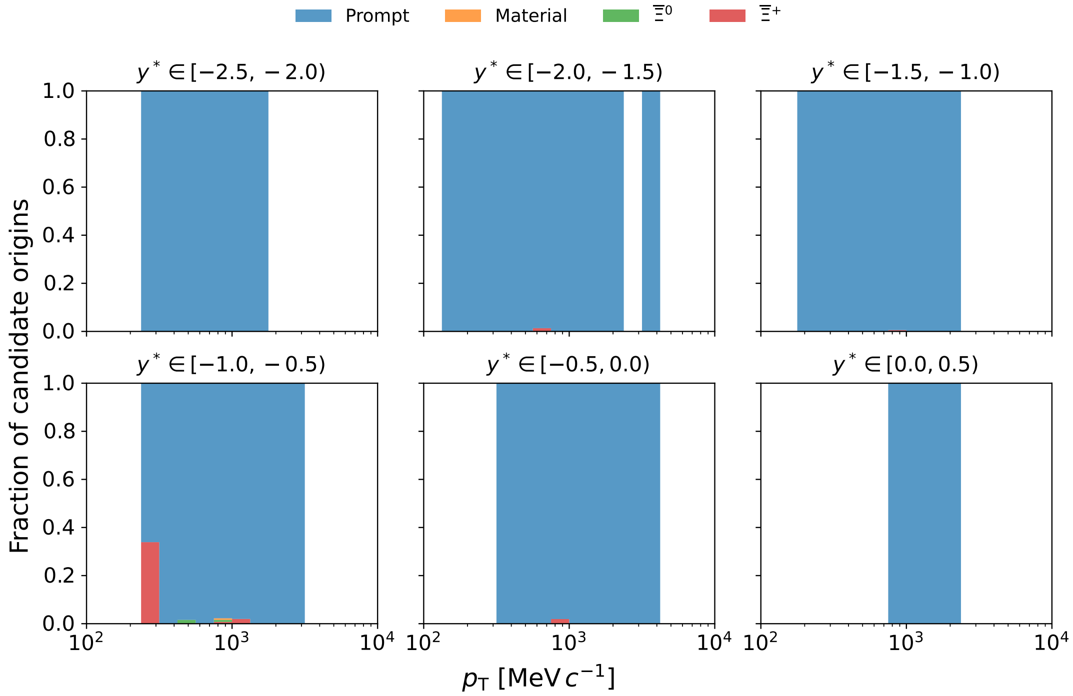


Figure 4.16 – Origins of $\overline{\Lambda}$ particles in the simulated *p*-He sample.

4.4.6 Total efficiency

Not every prompt V^0 particle produced in proton-nucleus collisions can be successfully reconstructed within this analysis. Some decay products may fall outside the geometric acceptance of the detector, for instance due to the magnetic field bending particle trajectories out of the sensible regions, into material of the beam-pipe, or detector itself. Additional losses arise from inactive detector regions, where traversing particles cannot be registered. Furthermore, the track-reconstruction algorithm itself may fail, particularly in high-occupancy events or when particles decay mid-flight, within the tracking system. The amount of observed V^0 particles is further reduced by the applied selection and mass veto. To account for these effects, the total efficiency is determined to correct the observed candidate count, as defined in Eq. (4.2).

The total efficiency can be determined using the simulated samples, which already includes the description of geometric acceptance, tracking and reconstruction efficiency. The total efficiency can be determined the number of generated V^0 particles and count the true, reconstructed V^0 particles that pass the selection criteria. The total efficiency is then given by

$$\epsilon_{\text{total}} = \frac{N_{\text{reco}}^{\text{selected}}}{N_{\text{gen}}}. \tag{4.14}$$

where N_{gen} is the number of generated prompt V^0 particles and $N_{\text{reco}}^{\text{selected}}$ is the number reconstructed prompt V^0 particles that passed the candidate selection. The total V^0 efficiency gained from the p-He simulated sample is illustrated in Fig. 4.17, in intervals of $y^* - p_{\text{T}}$.

The K_{S}^0 efficiency is consistently larger than the Λ and $\overline{\Lambda}$ efficiency. This cannot be caused by the BDT, as the efficiency is set to be 0.9 for each V^0 species. Therefore, this difference is likely mainly driven by the shorter lifetimes of K_{S}^0 particles compared to Λ and $\overline{\Lambda}$ particles, as well as the additional requirement applied on the impact parameter from Λ and $\overline{\Lambda}$ baryons. The efficiencies for Λ and $\overline{\Lambda}$ particles are approximately equal for the majority of bins. This is expected due to their particle-antiparticle relation, resulting in nearly identical kinematic distributions and decay topologies. Small differences can arise from the interaction of the decay products with the detector material, mainly the absorption and annihilation of antiprotons in the detector material. Further asymmetries can arise from the charged pions of the decays. When the Λ or $\overline{\Lambda}$ particle is produced at mid-rapidity, their corresponding pions may be bent out of detector acceptance, depending on the magnetic polarity. This effect, combined with the low statistics as indicated by their uncertainties, is most likely responsible for the observed asymmetry.

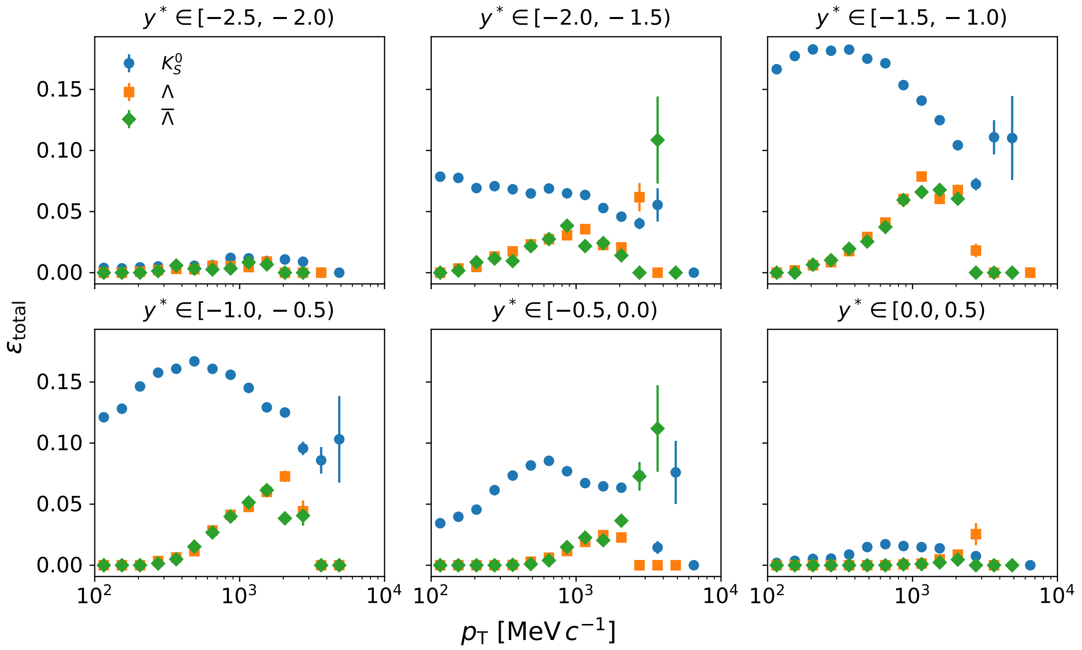


Figure 4.17 – Total efficiency for prompt K_S^0, Λ and $\overline{\Lambda}$ in intervals of $y^* - p_T$ for the p-He simulated sample.

Correction of the tracking efficiency

The total V^0 efficiency presented above already includes the tracking efficiency, which is not correctly simulated. To account for differences between the real and simulated samples, correction factors need to be applied. The correction factors are provided by the `Tracking Group` of the LHCb collaboration and are determined using $J/\psi \rightarrow \mu^+\mu^-$ decays in pp collisions [118]. The procedure in Ref. [118] describes the results for the 2010, 2011 and 2012 datasets, but the procedure for the 2016 and 2017 datasets is the same. The tracking efficiency correction is determined in proton-proton collisions, based on a *tag and probe* approach, and is provided in intervals of momentum and pseudorapidity in the laboratory frame. The decay is chosen due to its abundance and clean signature. Tag tracks are determined by selecting a long track, while probe tags are reconstructed using only hits in the T stations and the muon chambers. To keep the muon purity high, additional soft constraints are applied to both the tag and the probe track, based on the PID response and their kinematic and geometric features. Afterwards, two sets of J/ψ candidates are selected, based on the combination of either one tag and one probe track or two long-tracks. Soft constraints are again applied to reduce unwanted contributions, this time to the invariant di-muon mass and the geometric features of the reconstructed particles. The tracking efficiency is then determined by fitting the invariant mass distribution of both sets and then calculating the ratio of the results from the tag-and-probe sample to those from the two-long-track sample. The correction factor is subsequently calculated as the ratio of the efficiency in data over the efficiency in simulation. Intervals in which this method was not successful are assigned the correction factor 1.00 ± 0.05. An additional systematic uncertainty of 0.8 % is determined by varying the occupancy weighting of the simulated sample. The correction factors are illustrated for simulated samples corresponding to 2016 data-taking conditions in Fig. 4.18. The correction factors corresponding to 2017 data taking conditions are given in Appendix A.5.

The above described procedure does not factor in the loss of particles due to hadronic interactions, as it is performed using (anti)muons. For pions, the `Tracking Group` recommends an uncertainty of 1.4 % per track [118]. No uncertainty is given for protons, which is therefore estimated using the simulated samples. Just like (anti)muons, (anti)protons are long-lived charged particles that do not decay within the LHCb detector. As a result, the only difference in their tracking efficiency must stem from effects of hadronic interactions. Therefore, the hadronic proton loss, f_p, is estimated using the relation

$$f_p = 1 - \frac{\epsilon_p}{\epsilon_\mu}, \tag{4.15}$$

where ϵ_p is the tracking efficiency of protons and ϵ_μ the tracking efficiency of muons. The efficiencies are calculated based on Eq. (4.14). Due to the limited muon-statistics in the simulated sample, only a single factor integrated over all kinematic intervals is provided. For the p-He simulated sample, a hadronic loss of $f_p = 0.90 \pm 0.06$ is found for protons, while a hadronic loss of $f_{\bar{p}} = 0.69 \pm 0.05$ is found for antiprotons. For the p-Ne simulated sample, $f_p = 0.85 \pm 0.06$ and $f_{\bar{p}} = 0.65 \pm 0.04$ are found.

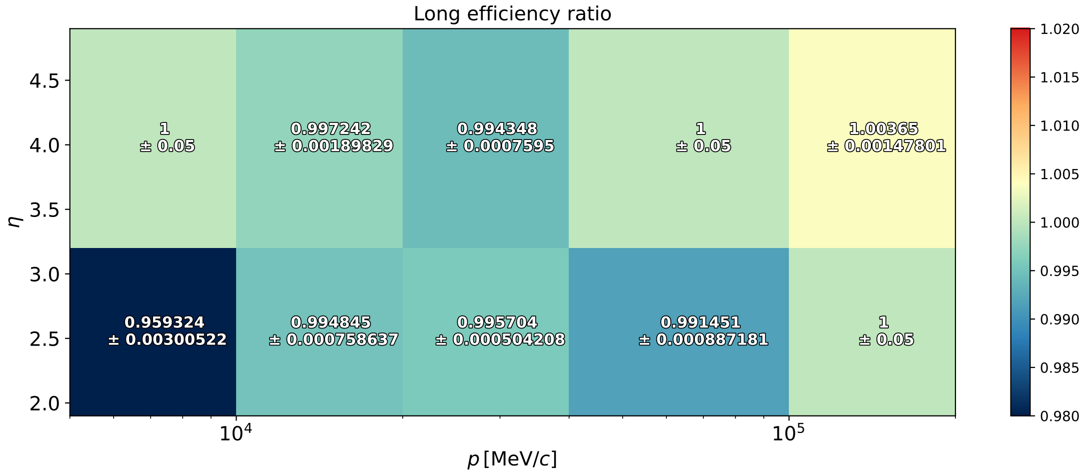


Figure 4.18 – Map of correction factors to account for discrepancies between real and simulated samples for 2016 data-taking conditions. The values are provided by the `Tracking Group` [123], using the procedure described in Ref. [118].

For each reconstructed V^0 in the simulated samples, the weight

$$w = f_1 \cdot f_2 \tag{4.16}$$

is applied, where f_1 and f_2 are the per-track correction factors. As the mapping does not cover the full phase-space studied in this analysis, the remaining intervals are assigned a correction factor of 1.00 ± 0.05, the same factor that is applied when the procedure described above failed. To account for variations across the different kinematic intervals between the provided mapping and the intervals employed in this analysis, a bootstrap approach is applied. In this approach, the correction factors are varied 100 times according to a Gaussian variation, with the standard deviation set to the uncertainty of the factors. The resulting correction factors, including their statistical and systematic uncertainties, are illustrated in Fig. 4.19, Fig. 4.20 and Fig. 4.21 for K_S^0, Λ and $\overline{\Lambda}$, respectively, for the p-He simulated sample. The correction factors for the p-Ne samples are presented in Appendix A.5. The distributions for K_S^0 show statistical uncertainties of up to 10 % for low p_T. As the correction factor is exactly one for these intervals, this indicates that both decay products are not covered by the kinematic intervals provided by the `Tracking Group`. The abrupt rise in the statistical uncertainty observed in high p_T is most likely linked to one of the decay products being in the kinematic interval where the method described above failed. This is further indicated by their corresponding distributions for the 2017 simulated samples, where these structures are not present. Similar trends are observed for the Λ and $\overline{\Lambda}$ distributions. However, the overall statistical uncertainty at low y^* is slightly lower than for K_S^0 mesons, remaining above 5 % but rarely reaching 10 %. In addition, the number of intervals in which a correction factor could be determined is smaller compared to K_S^0 mesons, reflecting differences in decay and track kinematics. For every V^0 particle, this correction constitutes one of the leading sources of overall uncertainty.

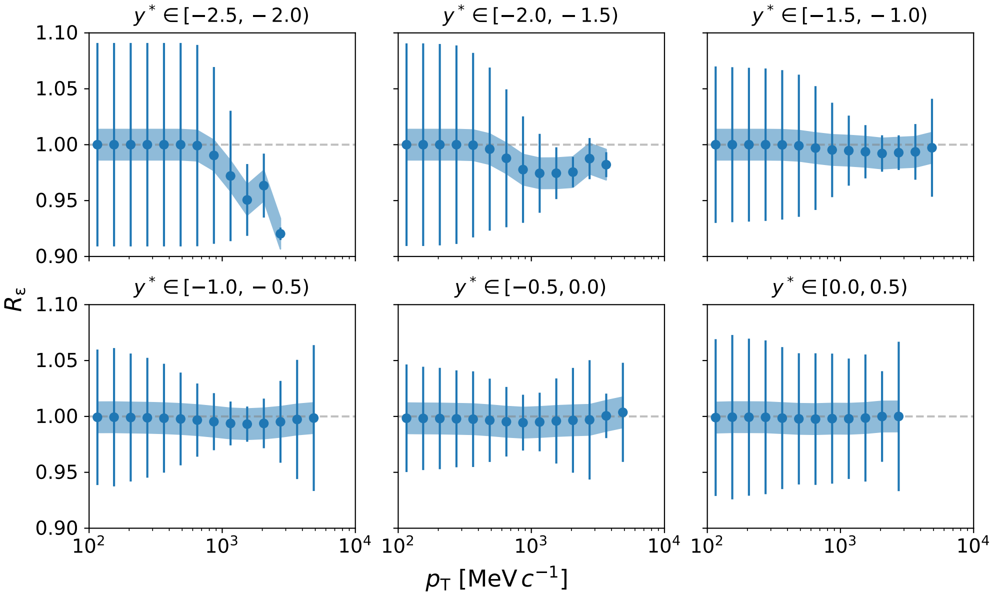


Figure 4.19 – Tracking correction for $K_S^0 \to \pi^+\pi^-$ decays for the p-He simulated sample. The bars and boxes represent the statistical and systematic uncertainty, respectively. The grey dashed line serves as a visual guide, indicating the value of 1.

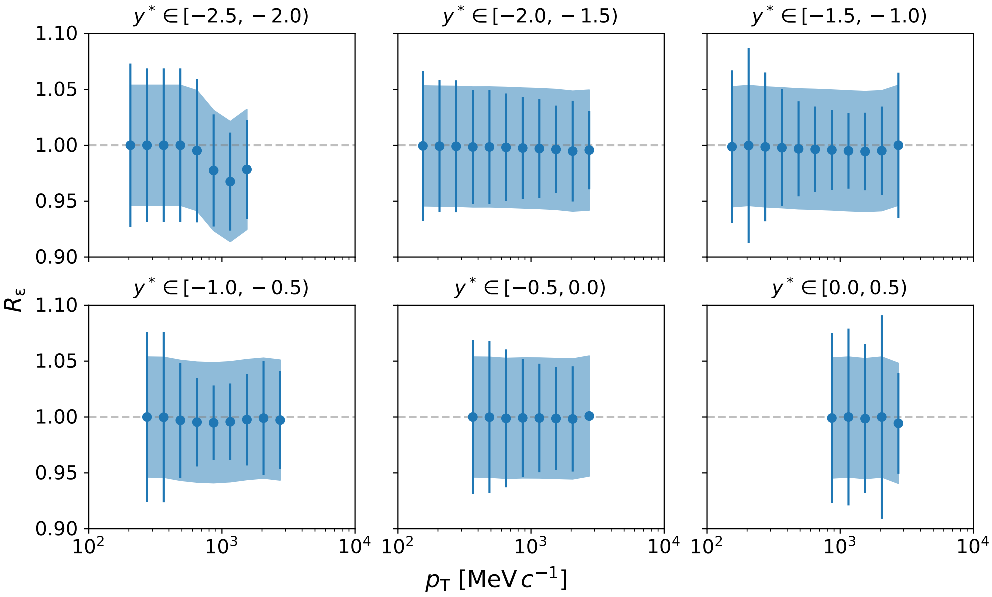


Figure 4.20 – Tracking correction for $\Lambda \to p\pi^-$ decays for the p-He simulated sample. The bars and boxes represent the statistical and systematic uncertainty, respectively. The grey dashed line serves as a visual guide, indicating the value of 1.

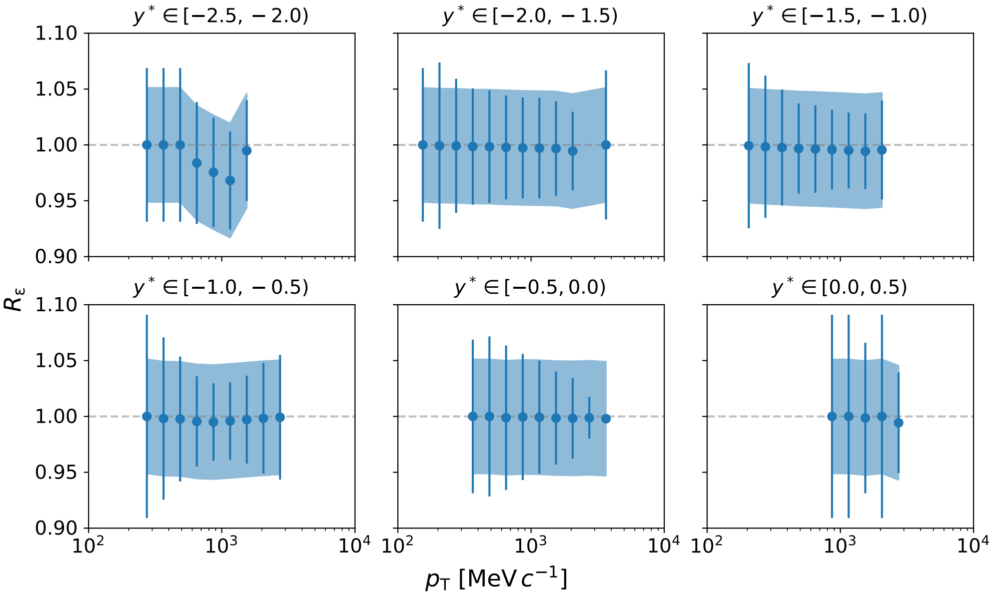


Figure 4.21 – Tracking correction for $\overline{\Lambda} \to \overline{p}\pi^+$ decays for the p-He simulated sample. The bars and boxes represent the statistical and systematic uncertainty, respectively. The grey dashed line serves as a visual guide, indicating the value of 1.

Decay time correction

As already mentioned in Sec. 4.4.1, most forward-going V^0 particles decay within LHCb detector as their mean lifetime is sufficiently short. However, their lifetime is long enough that a small fraction of them decay outside the detector volume. A feature of the LHCb simulation is that neutral particles not decaying within the LHCb acceptance are not saved. As a result, the number of generated particles used in Eq. (4.14) is reduced, thus overestimating the efficiency.

This bias is studied by reconstructing V^0 lifetimes. For an unstable particle with mean lifetime τ, the expected number of particles remaining as a function of time, $N(t)$, is exponentially decreasing and given by

$$N(t) = N_0 \cdot \exp\left(-\frac{t}{\tau}\right), \tag{4.17}$$

where N_0 is a scaling factor which corresponds to the number of promptly produced particles. The lifetime distribution can be reconstructed by calculating the flight distance of promptly produced V^0 particles. This is achieved using the vertex of origin, OV, where the V^0 is produced, and the decay vertex, DV, where it subsequently decays. Using both the V^0 mass and momentum, m_{V^0} and p_{V^0}, the decay time of the particle can then be calculated via

$$t = \frac{\|\vec{\mathrm{OV}} - \vec{\mathrm{DV}}\| \cdot m_{V^0}}{c \cdot p_{V^0}}. \tag{4.18}$$

The lifetime distribution is then determined via the decay time of each promptly produced V^0 particle. If V^0 particles decay exclusively within the detector acceptance, the resulting distribution matches the expected distribution in Eq. (4.17). Missing decays of generated V^0 particles, however, cause the reconstructed distribution to stop following the exponential law towards longer decay times. This effect is expected to increase as a function of momentum, as the generated V^0 particles take less time to leave the detector.

The correction factor, f, is determined by fitting the exponential decay law given in Eq. (4.17) to the reconstructed lifetime distribution. In this fitting procedure, the mean lifetime τ of the V^0 particle is kept fixed, thus only the scaling factor N_0 is extracted. As the cut-off in high-p_T could bias the fit, it is only performed using a reduced range in which the drop-off is not present. Two representative fits for generated K^0_S and Λ particles in the p-He simulated sample are presented Fig. 4.22. For both fits, comparisons to the fit using the full range are given as well. Using this fit, the correction factor f is determined via

$$f = \frac{A_\mathrm{Hist}}{A_\mathrm{Fit}}, \tag{4.19}$$

where A_Hist represents the area defined by the histogram, and A_Fit represents the area under the fitted curve.

It is observed that the cut-off is present in every kinematic interval, with the discrepancy starting earlier as a function of transverse momentum. The correction factor is—apart from statistical fluctuations—by definition smaller than one, and can be interpreted as the efficiency of a generated V^0 particle retained in the simulated sample. The efficiency for K_S^0 particles is approximately 98 % for the majority of intervals, going down to 92 % towards the edges of kinematic acceptance. The efficiencies of Λ and $\overline{\Lambda}$ are slightly lower, reaching 95 % in the majority of intervals and 89 % towards the kinematic edges. This is expected due to the increased average lifetime.

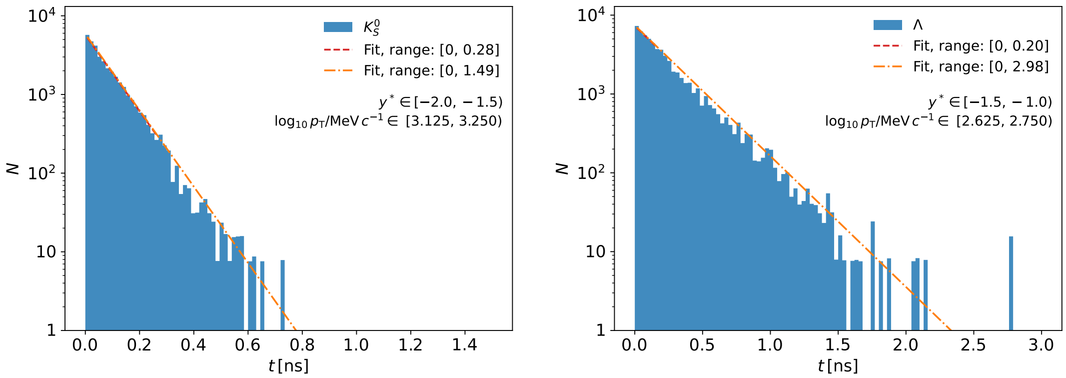


Figure 4.22 – Reconstructed decay-time distributions and fitted exponential decay laws for generated K_S^0 (left) and Λ (right) hadrons. The decay law is fitted in both a reduced and full interval.

4.5 Measurement of the prompt pion yield

In this section, the determination of the prompt charged pion yields is outlined. First, the selection of pion candidates using PID information is described. This is followed by the estimation of background contribution using the simulated samples. Following this, a proxy for fake tracks is determined, which is used in the subsequently presented template fits used to extract the prompt-pion count. Finally, the prompt-pion efficiency is determined.

4.5.1 Candidate selection

The majority of tracks stems from pions, which is why every track is a potential pion candidate. However, leading contributions can also stem from other charged particles, such as kaons, protons, electrons, muons, and their antiparticles. In addition, not all reconstructed tracks stem from promptly produced particles. Any long-lived particle decaying into at least one charged particle within the detector can generate additional tracks, which are associated to non-prompt particles. Also, due to ambiguities in the matching of track segments, the track sample is additionally contaminated by fake tracks, i.e. the incorrectly reconstructed particle trajectories not associated to a real particle passing through the detector.

While some of these contributions must be separated from the prompt charged pion component at a later step of the measurement, the first step is to reduce the contributions of other charged particles. This is done using the DLL variables provided by the PID system of the LHCb detector. In particular, three discriminating quantities are employed: the difference of the log-likelihood function with proton and pion hypotheses, $\mathrm{DLL}_{p\pi}$, kaon and pion hypotheses, $\mathrm{DLL}_{K\pi}$, and electron and pion hypotheses, $\mathrm{DLL}_{e\pi}$. These variables are smaller than zero if the track is more likely to stem from a pion and greater than zero if the track is more likely to stem from the respective other particle. Entries of zero indicate that the two particle hypotheses are equally likely or that the delta log likelihood could not be calculated. The latter is often the case for fake tracks. To ensure a high pion purity while keeping sufficient statistics, a selection is applied by requiring $\mathrm{DLL}_{p\pi} \leq -2.5$, $\mathrm{DLL}_{K\pi} \leq -2.5$, and $\mathrm{DLL}_{e\pi} \leq 0$. For the remaining tracks, the pion-mass hypothesis is applied, and their four-momenta corrected according to the procedure described in Sec. 4.3.2. The corrected four-momenta are then used in the subsequent analysis steps.

4.5.2 Candidate composition

The selection applied in Sec. 4.5.1 is expected to result in a sample with high pion-purity. The effectiveness of this selection is studied using the simulated samples. This is done using truth-matching, a procedure in which reconstructed tracks are matched with their corresponding, simulated particles. This procedure allows each track to be either identified as fake track, if no match is found, or correctly associated with a generated particle. The candidate composition of tracks with negative charge, before applying the selection, is presented in Fig. 4.23, for the p-He simulated sample in intervals of y^* and p_T. Only the relative contribution of associated particles, or lack of, to reconstructed long tracks, are shown. It does not contain any information about the origin, i.e. whether it was promptly produced or if it is the result of a material interaction.

As expected, the majority of long tracks stem from pions. Towards the edges of the kinematic acceptance, the majority of long tracks are fakes. This is expected as the amount of sensors decreases towards the edge of the geometric acceptance of the detectors, which results in fewer points for the tracking algorithm to fit, making fakes more likely. The remaining, dominant contributions stem from antiprotons, kaons and electrons. For these contributions, a p_T dependence is observed; the electron contribution is high for low p_T and decreasing towards higher p_T, while the antiproton and kaon distribution rise slightly towards higher p_T. The decreasing relative number of electrons is likely related to photon conversions, from which a majority of electrons are produced, as these are expected to happen less frequently for increasing p_T. The distributions for positively charged tracks shows the same trends and are thus not presented separately.

The candidate composition for negatively charged tracks, after the selection is applied, is presented in Fig. 4.24. The background contributions have been significantly reduced. The remaining leading background now comes from fake tracks, whose

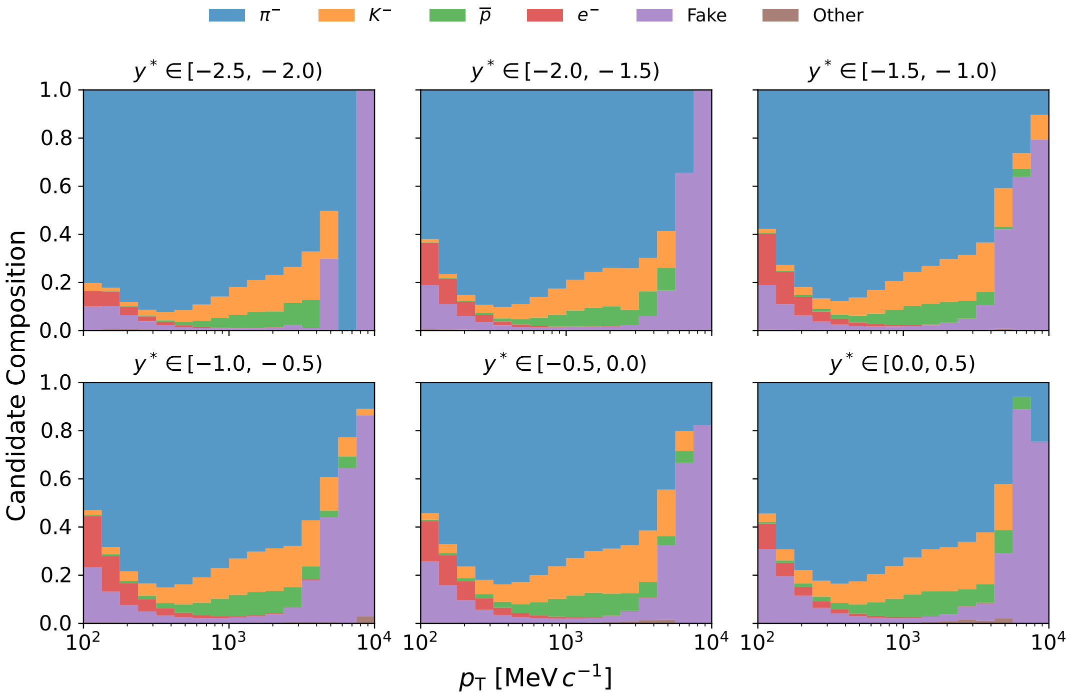


Figure 4.23 – π^- candidate composition in the p-He simulated sample in intervals of y^* and p_T, before the selection is applied.

relative contribution reaches its maximum at low p_T and high y^*, but falls significantly as p_T increases. No significant differences are observed in the corresponding π^+ distributions.

The prompt contribution from other particles needs to be investigated as it needs to be subtracted from the final prompt pion count according to Eq. (4.2). The same truth-matching approach as described above is used again, now only for prompt particles. The prompt π^- candidate composition is illustrated in Fig. 4.25 for the p-He simulated sample, in intervals of y^* and p_T. Again, no significant differences are observed for the remaining systems. Since the prompt π^- purity is high for the all kinematic intervals, with only a small prompt kaon contribution visible towards the edge of kinematic acceptance, these backgrounds are not investigated further. Instead, a 50 % systematic uncertainty is applied on their relative contribution.

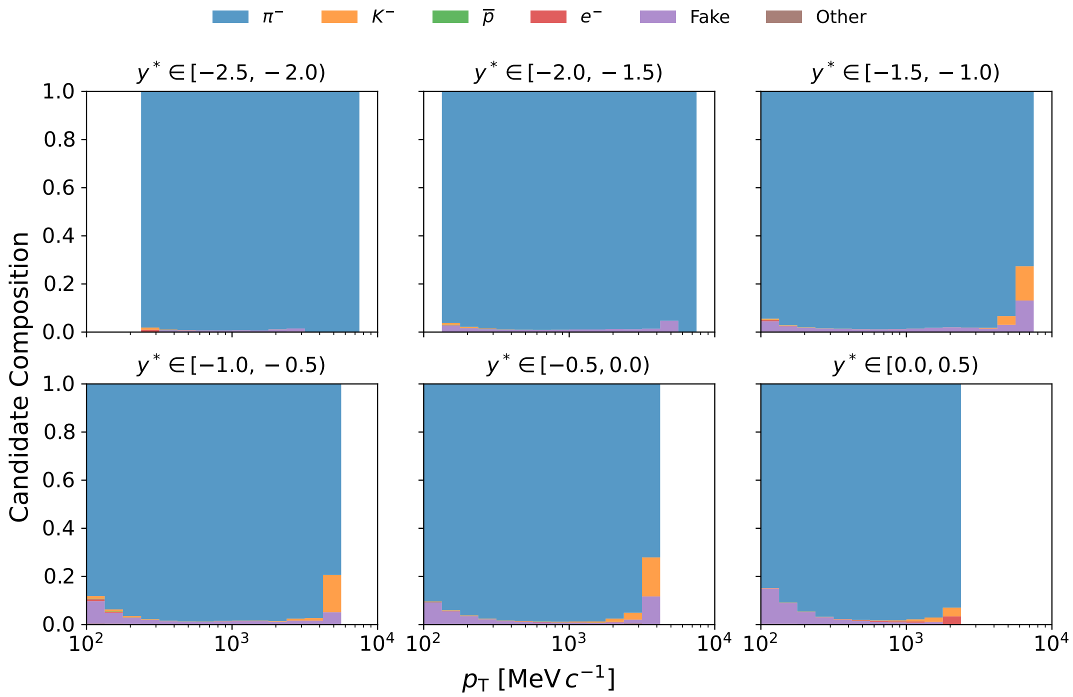


Figure 4.24 – π^- candidate composition in the p-He simulated sample in intervals of y^* and p_T, after the selection is applied.

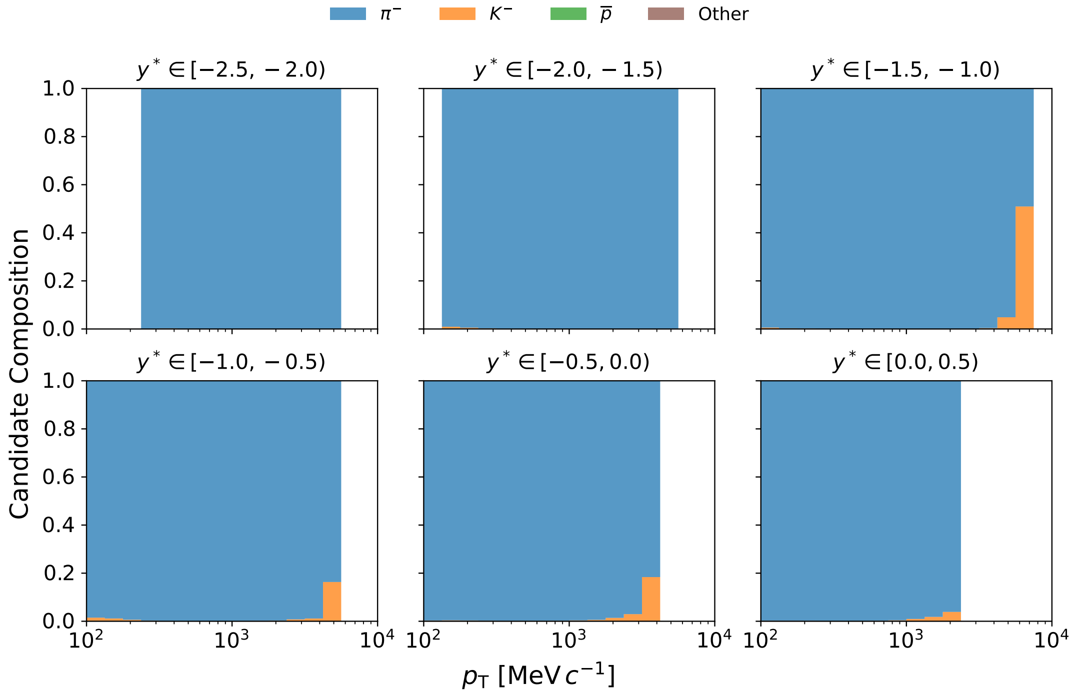


Figure 4.25 – Prompt π^- candidate composition in the p-He simulated sample in intervals of y^* and p_T, after the selection is applied.

4.5.3 Fake-track proxy

As discussed in the previous section, most background sources have already been significantly reduced. However, the remaining fake-track contribution can still bias the following measurement steps if left unconstrained. Due to this, a fake proxy is determined. Generally speaking, a proxy, $\mathcal{P}_k$, can be anything measurable which is proportional to the wanted quantity, n_k: $\mathcal{P}_k \sim n_k$. As this relation is equally valid for both real and simulated samples, it is possible to determine a ratio, R_k, which can be used to estimate the count of the real quantity in data using simulation:

$$R_k = \frac{\mathcal{P}_k}{\mathcal{P}_{\text{k, sim}}} = \frac{n_k}{n_{\text{k, sim}}} \tag{4.20}$$

$$\iff n_k = R_k \cdot n_{\text{k, sim}}. \tag{4.21}$$

A suitable fake proxy is a quantity with a fake purity as high as possible. For this purpose, the number of long tracks as a function of the fake probability, P_{fake}, is studied. Unfortunately, a cut on the fake purity was already applied during the online selection, leaving only the interval of $P_{\text{fake}} \in [0.0, 0.4]$ to work with. However, this window is sufficiently large for the purpose of this study. The fake purity can be determined using the simulated sample by calculating the figure of merit (FOM)

$$\text{FOM} = \frac{N_{\text{fake}}}{N_{\text{fake}} + N_{\text{real}}}, \tag{4.22}$$

where in this case, N_{fake} is the number of fake-tracks and N_{real} is the number of real tracks. The figure of merit is calculated for each kinematic interval within the two systems studied, as a function of P_{fake}. For this, the fake probability is divided into eight bins of equal width, with $\Delta P_{\text{fake}} = 0.05$. Example plots for the p-He simulated sample are presented in Fig. 4.26, in intervals of y^* and p_{T}.

The fake ratio, R_{Fake}, is determined using the number of long tracks with $P_{\text{fake}} \in [0.35, 0.40]$ per kinematic interval. Using the occupancy-reweighted simulated sample, the fake ratio can be estimated via

$$R_{\text{Fake}} = \frac{n^{\text{data}}_{\text{tracks}} - n^{\text{sig, sim}}_{\text{tracks}}}{n^{\text{fake, sim}}_{\text{tracks}}}, \tag{4.23}$$

where $n^{\text{data}}_{\text{tracks}}$ is the number of long tracks in the real sample, $n^{\text{sig, sim}}_{\text{tracks}}$ is the number of reconstructed long tracks in the simulated sample that are associated to generated particles, and $n^{\text{fake, sim}}_{\text{tracks}}$ is the number of fake tracks in the simulated sample. Due to the limited statistics of long tracks with high fake probability present after the selection applied in Sec. 4.5.1, the ratio is not determined separately for each charge. Instead, a single fake ratio is determined and later used for both charges. A systematic variation is performed by calculating the ratio again according to Eq. (4.23), but this time for long tracks with $P_{\text{fake}} \in [0.30, 0.35)$. The systematic uncertainty is then estimated by

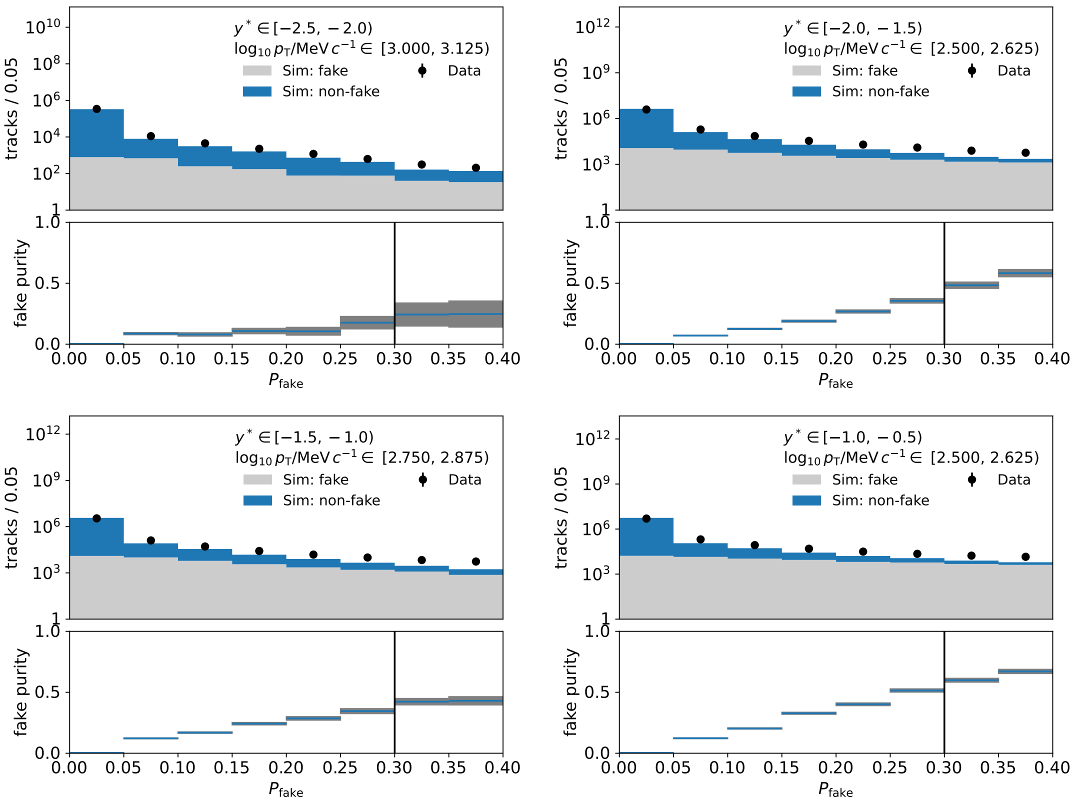


Figure 4.26 – Representative P_{fake} distributions in different $y^* - p_{\mathrm{T}}$ intervals for the p-He samples. Shown are the number of candidate tracks in the data sample (point) in the occupancy-weighted simulated sample (stacked histogram). The stacked histograms show the amount of fakes (grey) and matched tracks (blue). The lower plot in each example shows the fake purity. A vertical line is drawn at $P_{\mathrm{fake}} = 0.3$ serves as visual guidance, as the bins to the right of it will be used to determine the fake proxy and its systematic uncertainty.

calculating the absolute value of the difference of these two ratios. The resulting fake ratio for the p-He simulated sample is illustrated in Fig. 4.27, in intervals of $y^* - p_{\mathrm{T}}$. The proxy ratio is greater than unity in almost all kinematic intervals, indicating that the number of fake tracks is underestimated in simulation. In the first two y^* intervals, the proxy ratio fluctuates a lot as a function of p_{T}, reaching values of up to 10. These fluctuations, paired with high statistical and systematic uncertainties, suggest that is effect is caused by the limited amount of long tracks in these intervals, either in the real or simulated sample. These fluctuations and high uncertainties are however not a problem, as they will only serve as constraints in the following analysis steps. The distributions become smoother for the following y^* intervals, while also going towards unity, with the uncertainties decreasing as well. The amount of fake-tracks appears to decrease around a transverse momentum of 10^3 MeV/c, as indicated by the increasing statistical and systematic uncertainties and the subsequent absence of entries. Since the fake ratio is not determined separately for tracks associated with a positive or negative charge, the same ratio is used for π^+ and π^-, with a correlation of 100 %.

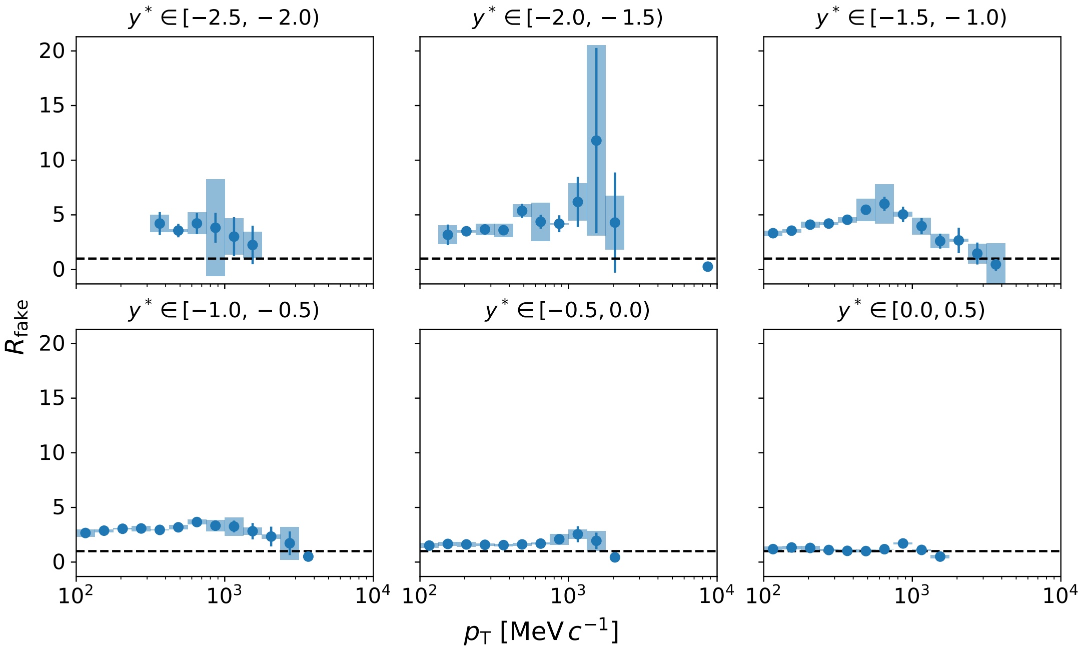


Figure 4.27 – Proxy ratio for fake tracks in intervals of y^* and p_T for the p-He sample. The bars and boxes represent the statistical and systematic uncertainty, respectively. The black dashed line serves as a visual guide, indicating the value of 1.

4.5.4 Signal yield extraction via template fits

As discussed previously, the pion purity in the remaining sample is very high. However, the origin of these pions has not been investigated yet. Consequently, the sample still contains contributions from non-prompt pions, originating from decays of long-lived particles and material interactions. In addition, contributions from fake-tracks persist and have to be taken into account.

In order to distinguish between prompt and non-prompt pions, a discriminating variable is required that is sensitive to the production origin of the track. A suitable variable for this purpose is the impact parameter: prompt pions, produced in proximity to the primary vertex, are characterized by a small impact parameter, while non-prompt pions extend to larger values, as they do not point back to the primary vertex. While the impact parameter provides strong discrimination power between prompt and non-prompt pions, simply applying a cut would reduce the available statistics and would not exploit the full available information. A more robust approach is therefore to perform a template fit. Using the simulated sample, the impact parameter distribution is modelled as superposition of the prompt, non-prompt and fake component. In a subsequent step, these components are scaled to unity and fitted on data, thus serving as a template. The fitted scaling factors then provide the count of the respective components in data. However, even though the fake contribution is small, it is observed to peak under the signal peak for certain kinematic intervals. This could lead to huge uncertainties, as the templates can then be chosen arbitrarily, resulting in the respective scaling factors being over- or underestimated in an attempt to match the data. Therefore, the fake component is restricted using the fake-proxy ratio determined in Sec. 4.5.3. The scaling factor is constrained by the product of the number of simulated fakes and the ratio; the lower and upper bounds are set by the propagated total uncertainties. In the majority of kinematic intervals, the prompt component peaks around an impact parameter of 0.1 mm, with the fake+non-prompt component starting to rise at around 1 mm. However, the templates do not perfectly match the data which, together with the small statistical uncertainties in data, results in a high χ^2/ndf. This is a common problem when working with template fits and not investigated further.

To evaluate potential modelling biases, a systematic variation of the template fit is performed. In this approach, the prompt peak is parametrized via a Double Sided Crystal Ball function [124] instead, while the non-prompt and fake components are fixed to their previously fitted templates. A Double Sided Crystal Ball function consists of a Gaussian core and two power-law tails. The Gaussian core is described using a normalization factor, N, the mean value μ, and the width parameters, σ_L and σ_R, of the left and right side of the Gaussian component. The power-law tails are described by two variables each: the exponent of the power-law tail, m_L or m_R, and the location of the transition to a power-law tail in standard deviation from the mean to the right,α_R, or left, α_L. To increase numerical stability, the Double Sided Crystal Ball function is not fitted on the IP, but on $\log_{10}$ IP/mm instead. To ensure convergence in every kinematic interval, soft constraints are applied to the parameters. The starting parameters are obtained from the true prompt distribution in the simulation. Example plots of both the pure template fit and the systematic variation for π^- particles are presented in Fig. 4.28 for chosen kinematic intervals in the p-He dataset.

The systematic uncertainty is then calculated as the absolute difference of the prompt pion counts obtained by the two methods. The resulting π^- counts, including their statistical and systematic uncertainty, are presented in Fig. 4.29, in intervals of y^* and p_T for the p-He dataset. The systematic uncertainty is on the sub-percent level for the majority of intervals, only reaching percent-level in intervals limited by statistics, in which the templates do not describe the data. The corresponding distributions for positively charged pions are similar and are thus not presented separately.

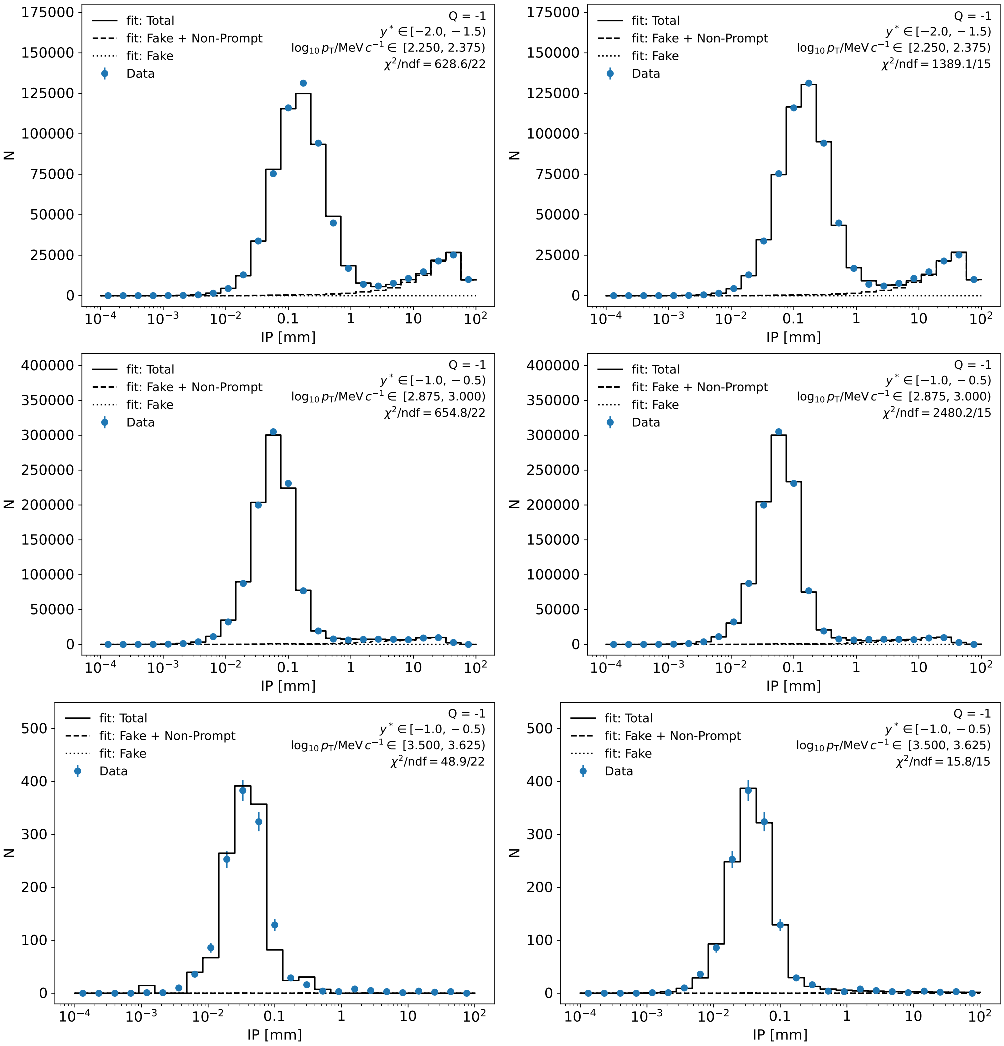


Figure 4.28 – Prompt π^- template fits for the p-He samples. The left column shows template fits for chosen $y^* - p_\mathrm{T}$ intervals, while the right column shows the corresponding systematic variations. The blue points represent the data, while the solid line represents the total fit model. The dashed and dotted line show the prompt+fake and fake components, respectively.

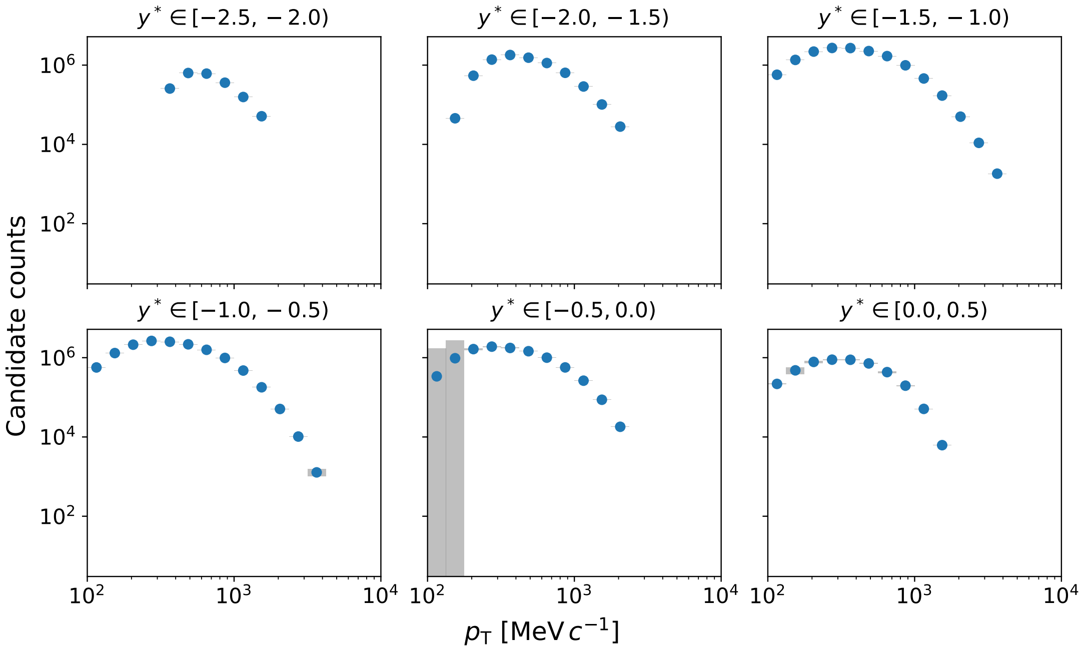


Figure 4.29 – Estimated prompt π^- counts for the p-He data sample in intervals y^* and p_T. The bars and boxes represent the statistical and systematic uncertainty, respectively.

4.5.5 Total efficiency

In this section, the total efficiency to observe prompt pions is discussed. Similarly to the prompt V^0 efficiency, the total prompt pion efficiency is the product of the tracking efficiency, i.e. the efficiency of a prompt pion to be reconstructed as a track, and the efficiency of a prompt pion to pass the applied selection, the PID efficiency. The determination of these efficiencies is presented in the following.

Prompt tracking efficiency

The prompt pion tracking efficiency is calculated using the reweighted simulated samples for each kinematic interval. This is done according to Eq. (4.14), where in this case, N_{gen} is the number of generated prompt pions, while $N_{\text{reco}}^{\text{selected}}$ is the number of tracks that can be associated to a prompt pion. It should be noted that no selection other than the event selection discussed in Sec. 4.3.1 is applied at this stage. Although it is in principle possible to combine the tracking efficiency directly with the PID efficiency by applying the DLL cuts, the PID efficiency is instead determined using data-driven methods, thus yielding a more accurate result. The prompt tracking efficiencies determined using the p-He simulated sample are illustrated in Fig. 4.30, in intervals of y^* and p_T. The efficiency multiplied by the pion charge is presented, allowing to display both the prompt π^+ and π^- efficiencies in the same plot without the points overlapping. Similarly to the efficiencies presented in Sec. 4.4.6, the prompt pion efficiencies are the lowest for low and high y^*, mostly only reaching efficiencies up to 0.5. The efficiencies are then slowly rising as a function of y^*, reaching their maxima at $y^* \in [-1,5,-1.0)$ and $y^* \in [-1.0,-0.5)$ before they are decreasing again.

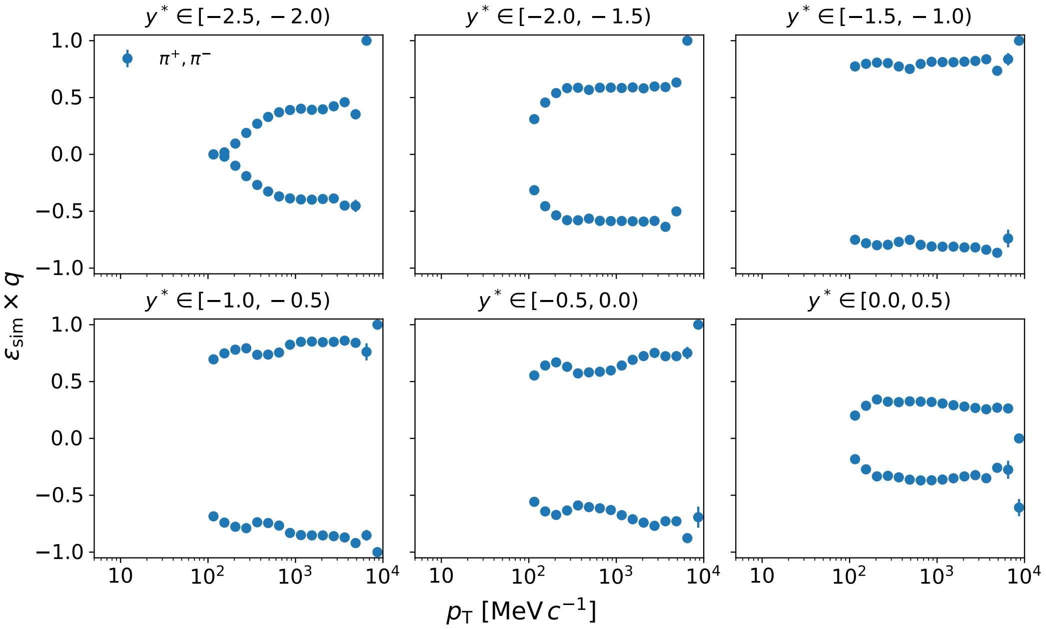


Figure 4.30 – Tracking efficiency of prompt pions in intervals of y^* and p_{T} for the p-He simulated sample. The product of the efficiency and the particle charge in units of the elementary charge is shown to separate values for oppositely charged particles.

Correction of the tracking efficiency

In analogy to the procedure described in Sec. 4.4.6, the prompt pion tracking efficiency must be corrected to account for mismatches between the data and simulated samples. The correction is again based on the correction factors provided by the `Tracking Group` [123].

Tracks associated with prompt pions are selected in the reweighted simulated sample and assigned a weight f, based on their kinematic properties. For pions falling outside the kinematic coverage of the provided correction table, the correction factor 1.00 ± 0.05 is applied, as it is the same correction employed by the `Tracking Group` for intervals in which no correction factor could be determined. The overall correction factor is obtained as the ratio of reweighted to the original number of tracks. Statistical uncertainties are determined by performing 100 Gaussian variations, where the $1\,\sigma$ intervals are set to the reported uncertainties. An overall systematic uncertainty of 0.8 % is applied. To account for hadronic interactions of pions, an additional systematic uncertainty of 1.4 % is considered, as recommended by the `Tracking Group` [118]. The correction factors for prompt π^- in the p-He simulated sample are presented in Fig. 4.31, in intervals of y^* and p_{T}. The statistical uncertainties are almost identical for prompt π^- and π^+, and reach up to 5 % for low p_{T}, indicating that these kinematic intervals are not covered by the `Tracking Group`. For the kinematic intervals $y^* \in [-1.5, -1.0)$, $y^* \in [-1.0, -0.5)$ and $y^* \in [-0.5, 0.0)$, the uncertainties, as a function of p_T, quickly rise up to 5 % before decreasing again. This is likely caused by the kinematic interval where the method described above failed. The same trends are observed for prompt π^+ particles and are thus not presented separately. The corresponding distribution obtained for the p-Ne samples is presented in Appendix A.5.

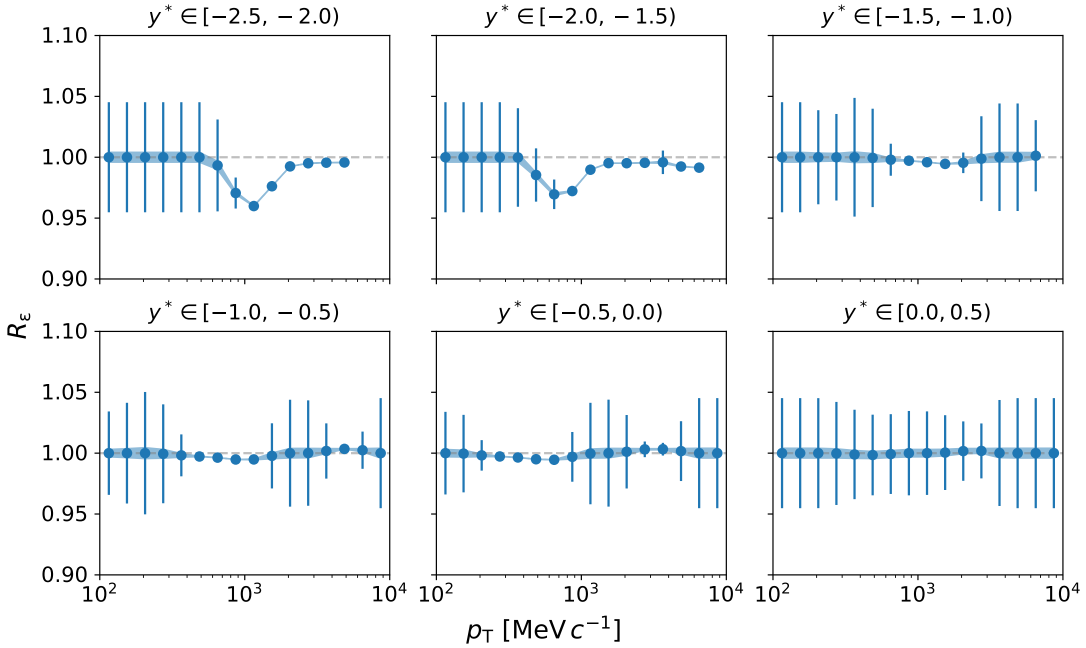


Figure 4.31 – Tracking correction for prompt π^- particles for the p-He simulated sample. The bars and boxes represent the statistical and systematic uncertainty, respectively. The grey dashed line serves as a visual guide, indicating the value of 1.

PID efficiency

Since the PID performance can differ between data and simulated sample, it is evaluated using the data samples. This is done using $K^0_\mathrm{S} \to \pi^+\pi^-$ decays as these provide a clean and well-defined source of pions; a clean $K^0_\mathrm{S} \to \pi^+\pi^-$ sample was already obtained through the application of a BDT in Sec. 4.4.2. It will therefore be re-used for this study.

The PID efficiency of a charged pion is determined by performing fits on the invariant $\pi^-\pi^+$ mass distribution, analogous to the fitting procedure described in Sec. 4.4.4. The fits are however not performed in the kinematic intervals of the K^0_S candidate, but in the kinematic intervals of the pion for which the PID efficiency is determined. The invariant-mass fits are performed two times, one time for all K^0_S candidates and one time for all K^0_S candidates where the pion candidate passes the DLL selection applied in Sec. 4.5.1. Example plots with and without the selection applied are presented in Fig. 4.32 for the p-He dataset.

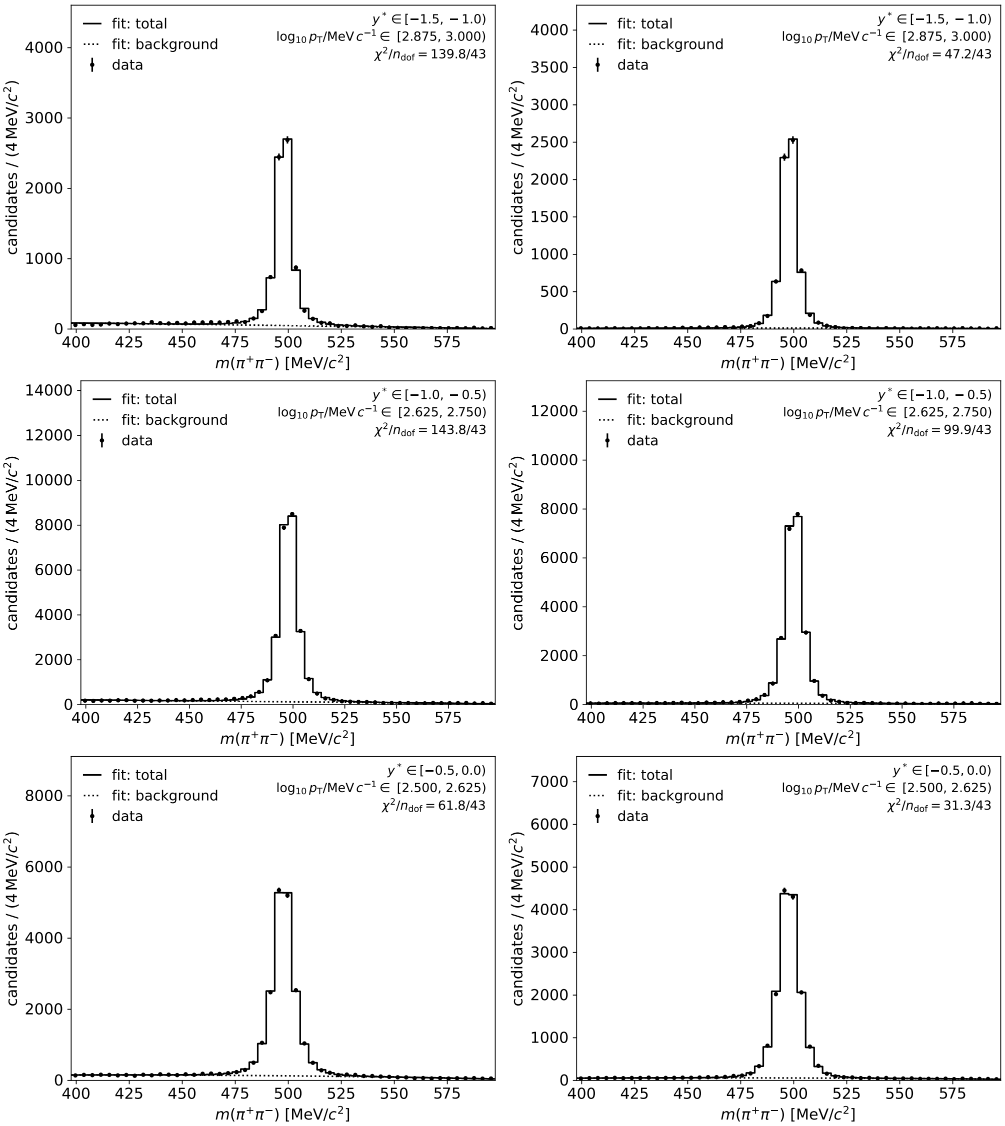


Figure 4.32 – Fitted invariant mass distributions obtained from the full K^0_S sample (left) and the K^0_S sample after applying the DLL selection (right) for the real p-He sample in chosen $y^* - p_\mathrm{T}$ intervals. The solid line shows the fitted complete model while the dashed line shows the background-component of the model.

The PID efficiency is then calculated from the fitted signal yields, $N_{\text{fit}}^{\text{all}}$ and $N_{\text{fit}}^{\text{cut}}$, obtained before and after applying the DLL selection, via

$$\epsilon_{\text{PID}}^{\text{fit}} = \frac{N_{\text{fit}}^{\text{cut}}}{N_{\text{fit}}^{\text{all}}}. \tag{4.24}$$

The statistical uncertainty is determined using a bootstrap approach. Every K_{S}^0 candidate is associated with an event, and every event is equally likely to occur. Therefore, the bootstrap is performed by drawing a sample of events and selecting every K_{S}^0 candidate associated to those events. In total, 100 bootstraps are performed, for each of which an efficiency is calculated. The statistical uncertainty is then given by the square root of the variance of these 100 efficiencies.

The systematic uncertainty is estimated using the simulated sample. Even if the true PID efficiency is not correctly described in the simulated sample, the method described above should be able to reproduce the simulated PID efficiency. Therefore, the true simulated PID efficiency, $\epsilon_{\text{PID}}^{\text{MC}}$, is determined by truth-matching tracks to prompt pions in the simulated sample. Using the number of truth-matched tracks before and after applying the PID selection, N_{MC}^{all} and $N_{\text{MC}}^{\text{cut}}$, the true PID efficiency of the simulated sample is determined as

$$\epsilon_{\text{PID}}^{\text{MC}} = \frac{N_{\text{MC}}^{\text{cut}}}{N_{\text{MC}}^{all}}. \tag{4.25}$$

Using the true simulated PID efficiency and the PID efficiency determined using the mass-fit method described above on the simulated sample, the systematic uncertainty is determined as the deviation from unity of the ratio $\epsilon_{\text{PID}}^{\text{fit, MC}}/\epsilon_{\text{PID}}^{\text{MC}}$. The resulting PID efficiencies for π^-, including their statistical and systematic uncertainty, are presented in Fig. 4.33, in intervals of y^* and p_{T} for the p-He data-sample. The statistical uncertainties are generally small, typically the order of 1 % in the majority of intervals, going up to 14 % towards the kinematic edges. The systematic uncertainty however is typically the order of 5 %, going up to 50 % towards the kinematic edges. This does not impact the measurement significantly, especially as they occur in regions where the overall uncertainties are generally large. Still, the PID efficiency is a leading source of systematic uncertainty overall. In individual intervals this uncertainty is even exceeding the uncertainties of the V^0 tracking correction. However, as the pions from $K_{\text{S}}^0 \to \pi^+\pi^-$ decays are not prompt, some small offsets within the kinematic intervals are expected. The prompt π^+ PID efficiency shows similar trends and is therefore not presented separately.

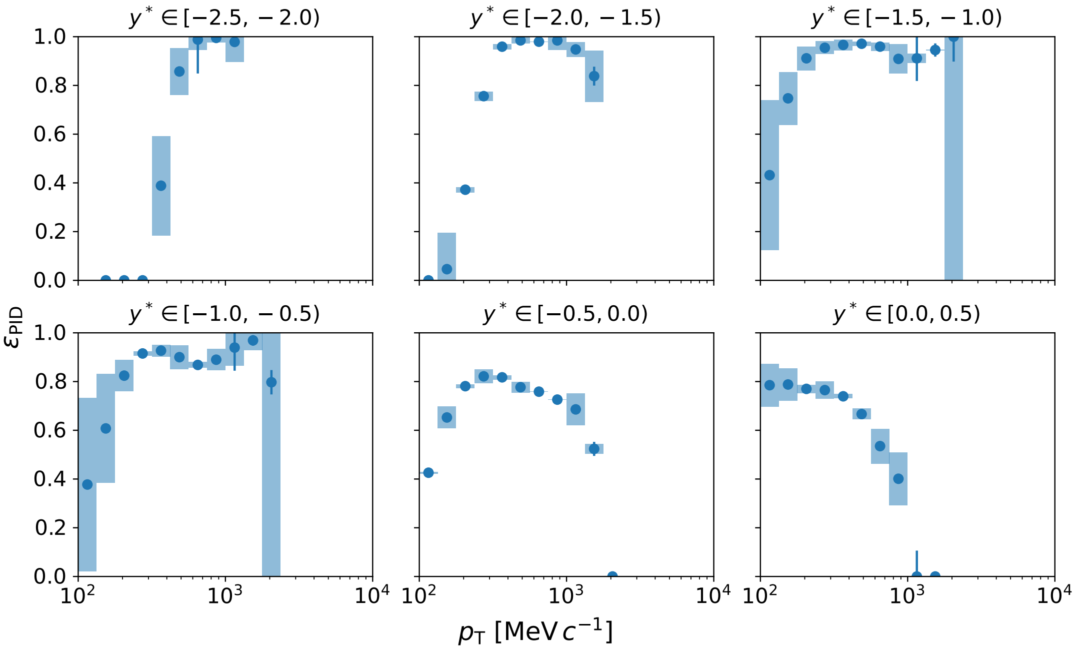


Figure 4.33 – PID efficiency obtained π^- from $K_S^0 \to \pi^+\pi^-$ decays in p-He collisions, in intervals of y^* and p_{T}. The bars and boxes represent the statistical and systematic uncertainty, respectively.

4.6 Results

The differential cross-section ratios are calculated according to Eq. (4.4). The full analysis workflow is applied for each dataset separately and presented here for both in intervals of y^* and p_{T}, and in intervals of p_{T} and $\langle N_{\mathrm{ch}}\rangle$. For this, the mapping of the number of long tracks to the average number of prompt, charged particles is applied as described in Sec. 4.3.4. In the following subsections, the propagated results for the studied ratios are presented. Kinematic intervals where no measurements could be made, either due lack of statistics or diverging uncertainties, are not presented. The results stemming from the p-He dataset are compared to generator predictions from `QGSJet-II-04` [54], `EPOS-LHC` [55] and `SIBYLL-2.3d` [56]. The generator predictions are obtained using `CHROMO` [125]; for each generator, 50 000 000 events are generated, from which the prompt particles are selected and the ratios calculated. Differences with respect to generator predictions are highlighted for both binning systems. Since the results obtained using the p-Ne samples are overall very similar to those obtained from the p-He samples, they are not discussed separately in this section, but presented in Appendix A.6 instead. As `SIBYLL-2.3d` does not yet support neon targets, p-Ne results are only compared to `QGSJet-II-04` and `EPOS-LHC`. To provide a comparison with the findings of the ALICE collaboration and interpret the results, the measurement performed in intervals of p_{T} and $\langle N_{\mathrm{ch}}\rangle$ is integrated in respect to p_{T}, which is separately discussed in Sec. 5.1.

4.6.1 Cross-section ratios of K_S^0 to charged pions

The obtained differential cross-section ratios for the p-He 2016 dataset, together with generator predictions, are presented in Fig. 4.34, in intervals of y^* and p_T. The measurement is dominated by systematic uncertainties. For the majority of intervals, the systematic uncertainties range from 5 to 12 %, while the statistical uncertainties remain below 2 %. For all intervals shown, both the measured ratios and generator predictions suggest an increased strangeness production as a function of transverse momentum. However, for the majority of kinematic intervals, the generators overestimate the ratio, while the overall observed trends align best with the predictions from `EPOS-LHC`. At low p_T, the predictions from `SIBYLL-2.3d` and `EPOS-LHC` are similar. However, `SIBYLL-2.3d` predicts a stronger increase around 1 GeV/c, followed by a flattening of the slope, after which it remains slightly offset relative to `EPOS-LHC`. The predictions from `QGSJet-II-04` start of higher at low p_T and continuously increase, with no noticeable reduction in slope towards higher p_T. For a few selected intervals however, the predictions from `EPOS-LHC` and `QGSJet-II-04` are compatible with data within the uncertainties of the measurement.

The differential cross-section ratios, in intervals of p_T and $\langle N_\mathrm{ch}\rangle$, are presented in Fig. 4.35. This measurement is again dominated by systematic uncertainties. For the majority of intervals, the systematic uncertainties range from 5 to 15 %, while the statistical uncertainties remain below 3 %. At low and high p_T, the generators reproduce the overall trend observed in data, although they systematically overestimate the ratio. In particular, at low p_T, the measured ratio is decreasing as a function of $\langle N_\mathrm{ch}\rangle$. This trend is reproduced by both `EPOS-LHC` and `SIBYLL-2.3d`, whose predictions are in close agreement with each other. In contrast, `QGSJet-II-04` shows a larger offset and predicts an approximately constant ratio. At mid p_T however, the overall trends of the event generators do not agree with the measurement. There, the measured cross-section ratio rises as a function of multiplicity before reaching a plateau. This behaviour is not reproduced by the generators, which all predict a slightly decreasing ratio, with minimal offsets to each other. At high p_T, all generators predict a slightly decreasing ratio as a function of multiplicity, while the measured ratio is approximately constant. Here, the predictions from `EPOS-LHC` are within the uncertainties of the measurement, while the remaining generators overestimate the ratio.

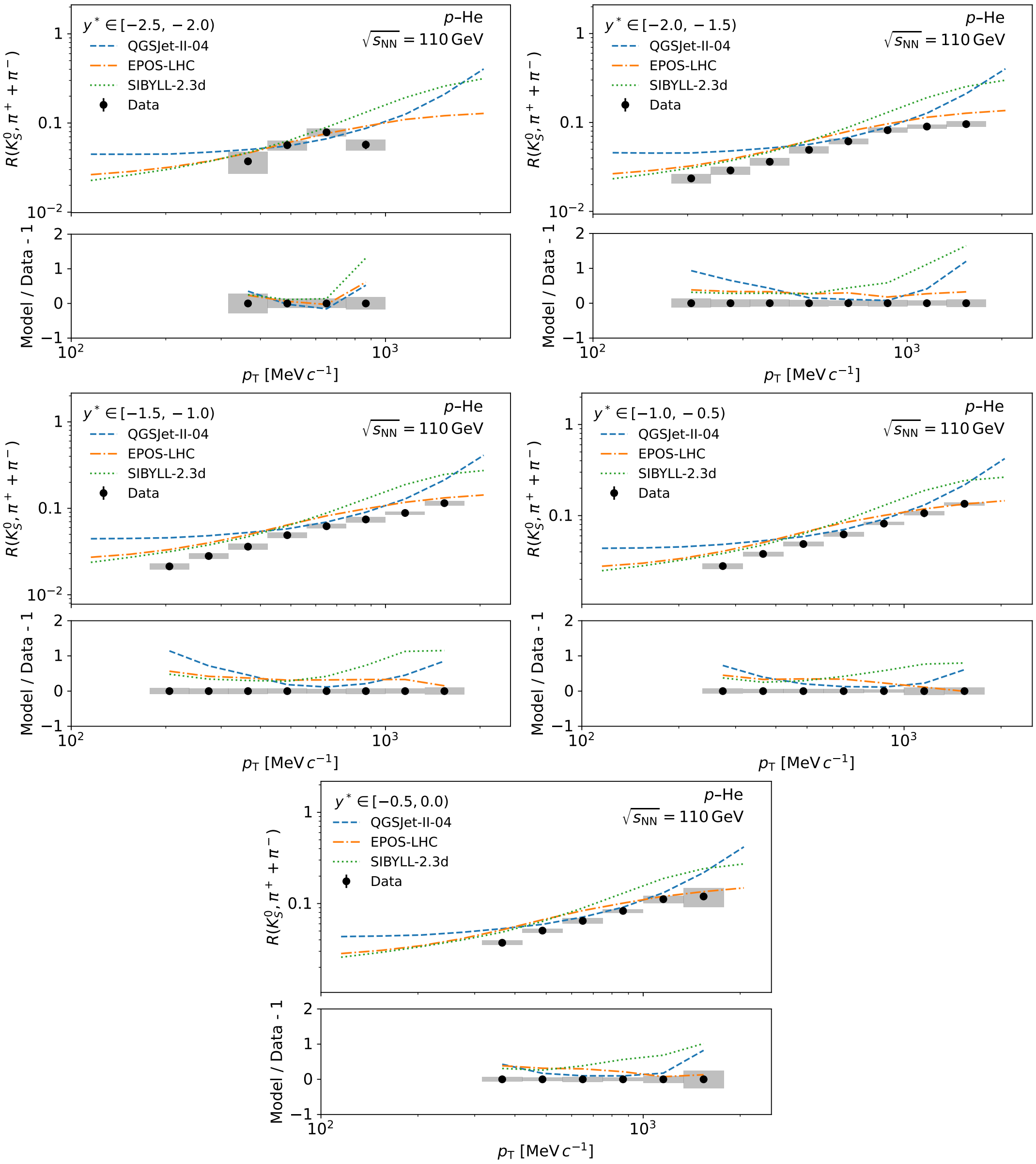


Figure 4.34 – Differential cross-section ratios of K_S^0 particles to charged pions, π^- and π^+, in intervals of y^* and p_T for the p-He 2016 dataset. The black points show the measured ratios, and coloured lines correspond to the predictions of different event generators. The bars and boxes represent the statistical and systematic uncertainty, respectively. The relative deviation of the models to the measured ratio is given in the lower part of each plot.

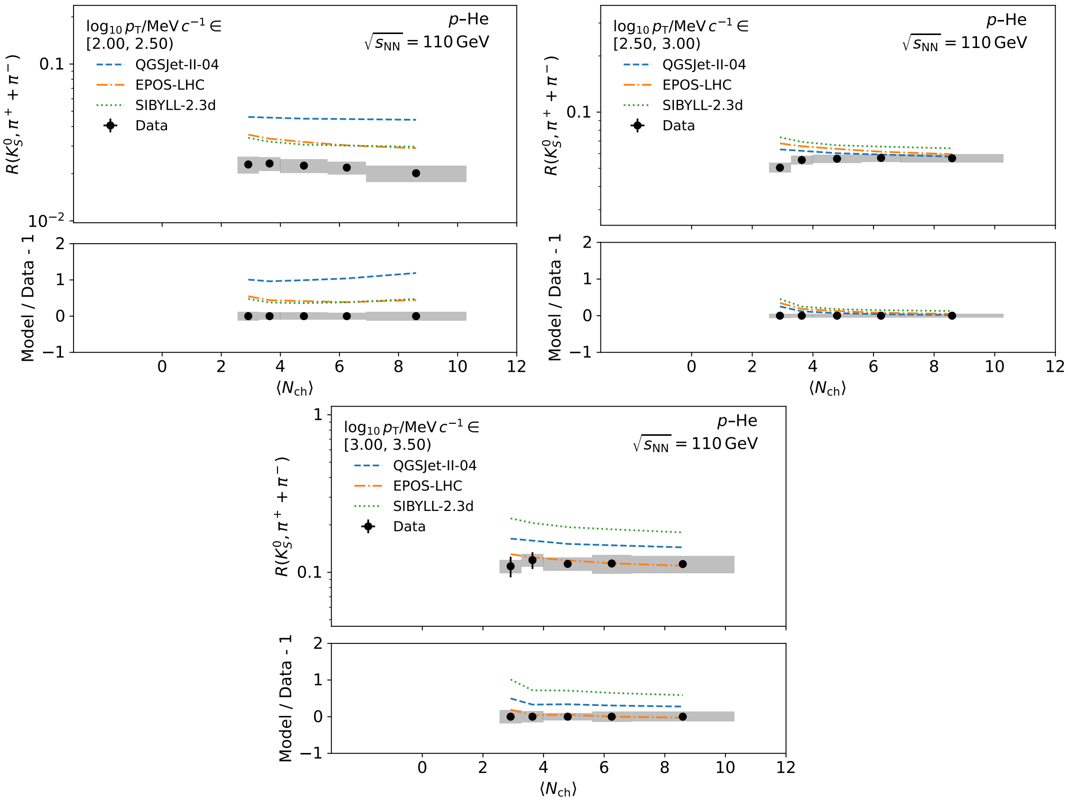


Figure 4.35 – Differential cross-section ratios of K_S^0 to charged pions, π^- and π^+, in intervals of p_T and $\langle N_{ch}\rangle$ for the p-He 2016 dataset. The black points show the measured ratios, and coloured lines correspond to the predictions of different event generators. The bars and boxes represent the statistical and systematic uncertainty, respectively. The relative deviation of the models to the measured ratio is given in the lower part of each plot.

4.6.2 Cross-section ratios of Λ and $\overline{\Lambda}$ to charged pions

In this section, the differential cross-section ratios $(\Lambda + \overline{\Lambda})/(\pi^+ + \pi^-)$, derived from the p-He dataset, are presented. The ratios in intervals in y^* and p_{T} are illustrated in Fig. 4.36. For this ratio, systematic uncertainties dominate the measurement, although statistical uncertainties remain non-negligible in the majority of intervals. For most intervals, the systematic uncertainties range from 6 to 17 %, with the uncertainties increasing towards the kinematic edges. The statistical uncertainties range between 3 to 11 % for the majority of intervals, again increasing towards the kinematic edges. The measured ratios increase as a function of transverse momentum, a trend replicated by the generators. Overall, the predicted ratios by the generators are similar, with `QGSJet-II-04` yielding slighter larger values than `SIBYLL-2.3d` and `EPOS-LHC`. The generated ratios mainly differ in their slope towards higher p_{T}, with `QGSJet-II-04` exhibiting the steepest increase, followed by `SIBYLL-2.3d` and then `EPOS-LHC`. Although the generators underestimate the ratio in the majority of intervals, the measured ratios towards high p_{T} are best described by `SIBYLL-2.3d`.

The cross-section ratios for intervals of p_{T} and $\langle N_{\mathrm{ch}} \rangle$ are illustrated in Fig. 4.37. The ratios are again systematically dominated, although the statistical uncertainty is non-negligible, especially towards low and high p_{T}. For most intervals, the systematic uncertainties range from 6 to 18 %, with the uncertainties increasing towards the kinematic edges. The statistical uncertainties range between 3 to 11 % for the majority of intervals, again increasing towards the kinematic edges. For low p_{T}, the measured ratio is decreasing as a function of multiplicity. This trend is predicted by both `QGSJet-II-04` and `EPOS-LHC`, while `SIBYLL-2.3d` predicts a slightly increasing ratios. Overall, the generator predictions are largely consistent with the measurement within uncertainties; however, small deviations are observed in the first and last interval. At mid p_{T}, the measured ratio first exhibits a slight increase as a function of multiplicity, followed by a subsequent decrease. A slightly increasing ratio is predicted by `EPOS-LHC` and `SIBYLL-2.3d`, but all generators consistently underestimate the ratio. Finally, at large p_{T}, the measured ratio shows a slight increase as a function of multiplicity. This trend is again predicted by `EPOS-LHC` and `SIBYLL-2.3d`, while only the latter lies within the uncertainties of the measurement. In contrast to this, `QGSJet-II-04` predicts a slightly decreasing ratio. However, due to the large uncertainties, its predictions lie within the uncertainties of the measurement towards high multiplicity.

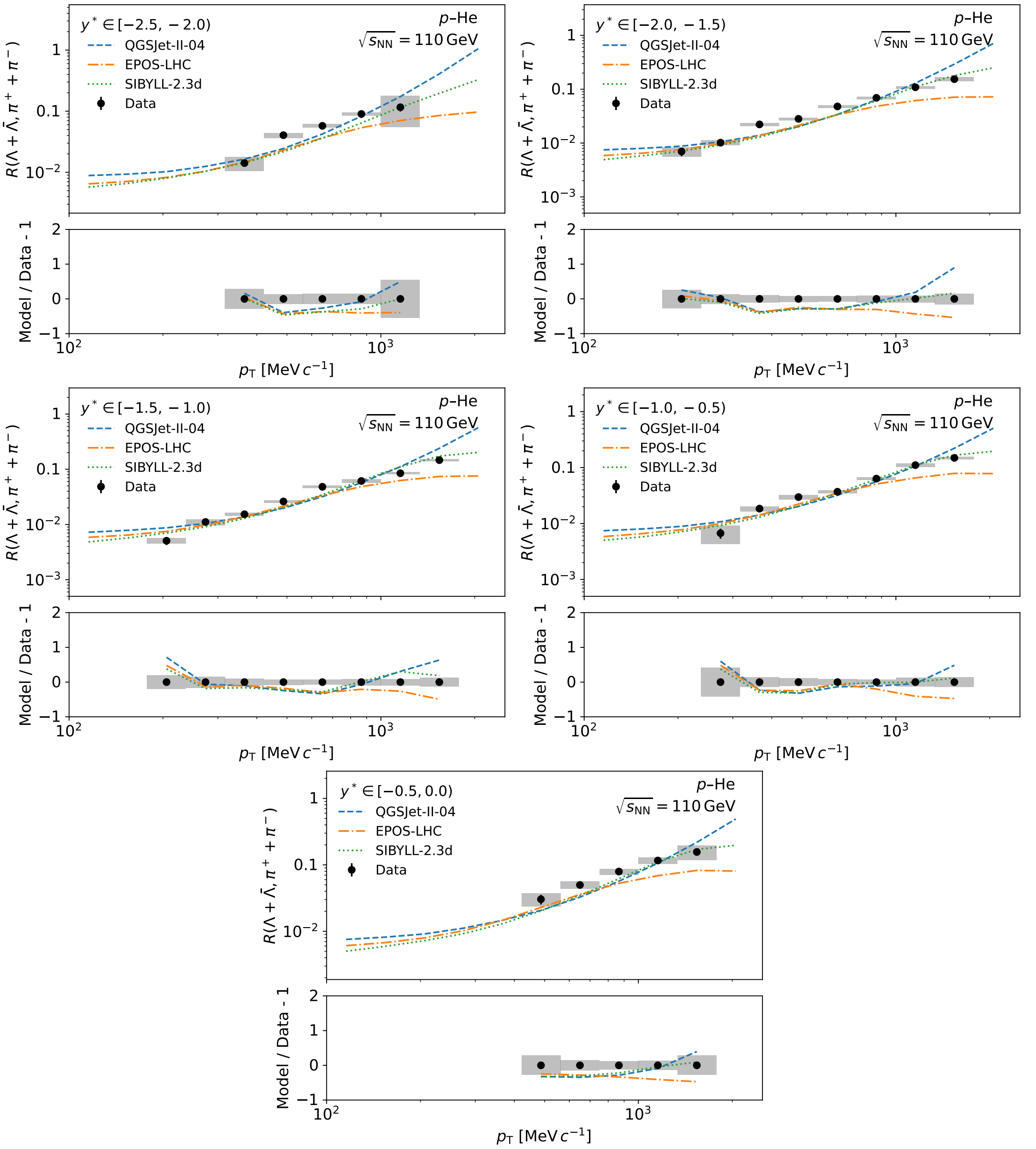


Figure 4.36 – Differential cross-section ratios of Λ and $\overline{\Lambda}$ particles to charged pions, π^- and π^+, in intervals of y^* and p_{T} for the p-He 2016 dataset. The black points show the measured ratios, and coloured lines correspond to the predictions of different event generators. The bars and boxes represent the statistical and systematic uncertainty, respectively. The relative deviation of the models to the measured ratio is given in the lower part of each plot.

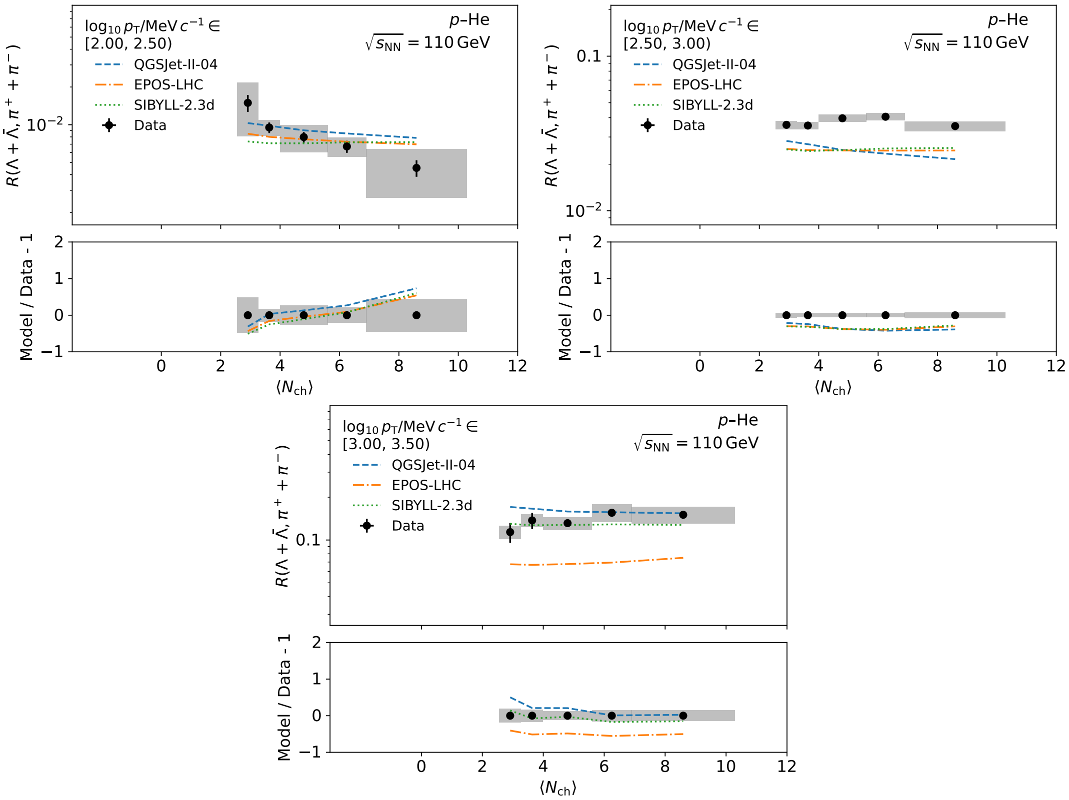


Figure 4.37 – Differential cross-section ratios of Λ and $\overline{\Lambda}$ particles to charged pions, π^- and π^+, in intervals of p_T and $\langle N_{ch} \rangle$ for the p-He 2016 dataset. The black points show the measured ratios, and coloured lines correspond to the predictions of different event generators. The bars and boxes represent the statistical and systematic uncertainty, respectively. The relative deviation of the models to the measured ratio is given in the lower part of each plot.

5 Towards a global picture

In the previous chapter, cross-section ratios of strange V^0 hadrons to charged pions have been presented in multiple collision systems and kinematic intervals. However, the interpretation of these results require them to be placed within a broader experimental context. A global picture is particularly essential for evaluating the universality of multiplicity dependent strangeness enhancement. In this chapter, the results are therefore compared to findings of the ALICE collaborations and additional work in this context is presented.

5.1 Comparison with ALICE

The results presented in Sec. 4.6 provide a double-differential measurement of strangeness production as a function of transverse momentum and multiplicity. This offers increased insights to the underlying particle production mechanism and is thus an important input to the generator tuning community. However, the measurement performed by the ALICE collaboration in Ref. [31], which motivates this work, is reported as a single-differential observable; strangeness production is measured as a function of multiplicity, integrated over transverse momentum. To provide a direct comparison to these measurements, and to place the results in the broader experimental context, a rough estimate of the p_{T}-integrated strangeness enhancement is provided. This is done by adding the efficiency corrected yields of the different p_{T} intervals before calculating the ratios in analogy to Eq. (4.4). It should be emphasized, however, that this is not a replacement for the integrated measurement. Individual p_{T} intervals suffer from large uncertainties, and correlations between the intervals are not taken into account. In addition, missing contributions from regions outside the used p_{T} intervals may also affect the results. The resulting single-differential cross-section ratios are therefore intended solely to provide a qualitative comparison to the published ALICE results.

The comparison is done using the publicly available HEPData entry corresponding to Ref. [31], which is available in Ref. [126]. As the measurement performed by the ALICE collaboration covers a different phase space—in terms of pseudorapidity and centre-of-mass energy—compared to the measurement presented by the author, the cross-section ratios may differ. However, only the shape of the cross-section ratio is relevant for the study of strangeness enhancement, not the absolute value. Therefore, to compare the trends as a function of multiplicity, both ALICE results and the presented, integrated results are normalized to their respective mean values. The resulting distributions are presented in Fig. 5.1. For both the p-He and p-Ne data samples, the estimated cross-section ratios of K^0_{S} particles to charged pions are within the uncertainties of the ALICE measurement. For both the p-He and p-Ne data samples, an initial rise is observed, which may hint at multiplicity dependent strangeness enhancement. However, the overall uncertainties are too large to draw any conclusions. In contrast to this, the estimated cross-section ratios of $\Lambda + \overline{\Lambda}$ particles to charged pions are not fully consistent with the ALICE results. This concerns mainly the p-He data sample, where the estimated ratios fluctuate as a function of multiplicity. Nevertheless, most data points remain compatible with the ALICE distributions, with noticeable deviations observed only towards low and high multiplicity. No clear evidence of multiplicity dependent strangeness enhancement is observed.

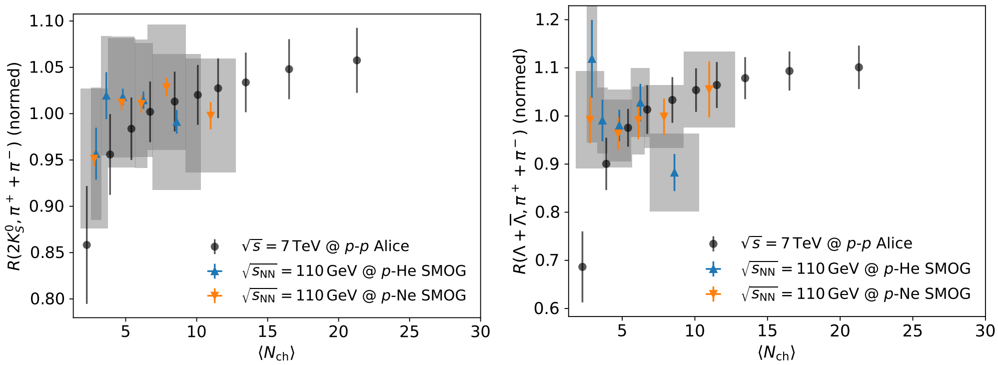


Figure 5.1 – Normalized, p_{T} integrated yields of $2K^0_{\mathrm{S}}$ (left) and $\Lambda + \overline{\Lambda}$ (right) to charged pions. The black markers represent the measurement performed by the ALICE collaboration [31, 126]. The blue and orange data points correspond to the estimated ratios for p-He and p-Ne collisions using SMOG, derived from the results presented in Sec. 4.6. In case of the data from the ALICE collaboration, the error bars represent the total uncertainties. For everything else, the error bar only refers to statistical uncertainty, while the grey box represents systematic uncertainty.

The comparison to the findings from the ALICE collaboration reflects the limitations of the current estimates, which include large systematic uncertainties. Differences may occur from the different coverages in pseudorapdity of the experiments. The ALICE experiment provides full tracking capabilities within $|\eta_{\mathrm{lab}}| < 0.9$ [83], while the LHCb experiment covers the forward region $\eta_{\mathrm{lab}} \in [2, 5]$. Accounting for the Lorentz-boost for the SMOG system in analogy to Eq. (4.9), the LHCb coverage in the centre-of-mass frame becomes approximately $\eta_{\mathrm{com}} \in [-2.8, 0.3]$. To visualize the expected differences between the performed measurement and that of the ALICE collaboration, normalized differential particle spectra as a function of rapidity in the centre-of-mass system are shown in Fig. 5.2, where the pseudorapidity-ranges fully covered by the experiments are explicitly highlighted. In doing so, pseudorapidity is used as a proxy for rapidity, which is valid for particles with momenta much larger than their masses ($p >> m$). The corresponding predicted ratios are presented in Fig. 5.3. As shown there, the ratios are expected to differ across the different collision systems and acceptances. Minor differences may also occur due to the slightly different measurement ranges used for the multiplicity-mapping; the ALICE collaboration employed $|y^*| < 0.5$, while the interval $y^* \in [-1.0, 0.0)$ is employed in the presented measurement.

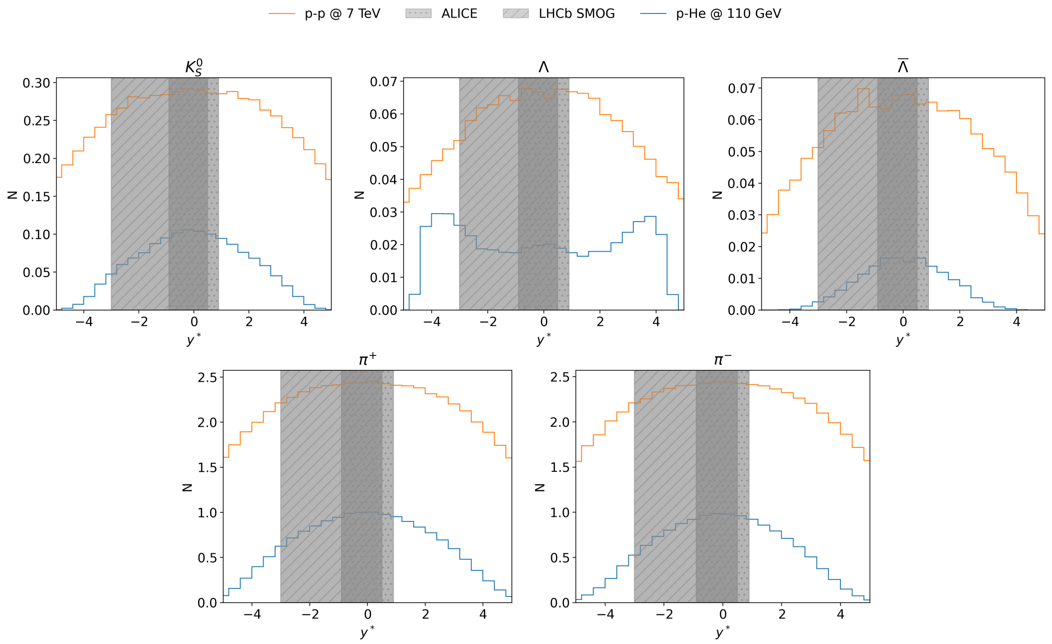


Figure 5.2 – Normalized differential particle spectra as a function of rapidity in the centre-of-mass system, as predicted by `EPOS-LHC`. The orange lines correspond the expected distributions for proton-proton collision at a centre-of-mass energy of 7 TeV, corresponding to the conditions in Ref. [31]. The blue curves show the distributions for proton-helium collisions at a nucleon-nucleon centre-of-mass energy of 110 GeV, corresponding the conditions of the presented measurements. The grey boxes correspond to the coverage of pseudorapidity in the centre-of-mass system of the respective experiments.

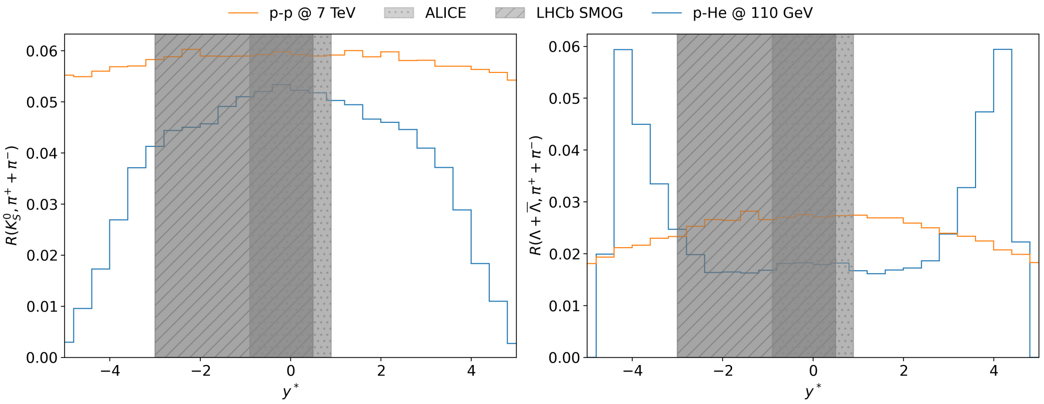


Figure 5.3 – Differential cross-section ratios as a function of rapidity in the centre-of-mass system, as predicted by `EPOS-LHC`. The orange lines correspond the expected distributions for proton-proton collision at a centre-of-mass energy of 7 TeV, corresponding to the conditions in Ref. [31]. The blue curves show the distributions for proton-helium collisions at a nucleon-nucleon centre-of-mass energy of 110 GeV, corresponding the conditions of the presented measurements. The grey boxes correspond to the coverage of pseudorapidity in the centre-of-mass system of the respective experiments.

Finally, the p_{T}-integrated measurement is compared with generator predictions to assess how well they reproduce the data, as presented in Fig. 5.4. For both p-He and p-Ne collisions, cross-section ratios involving K^0_{S} mesons are generally overestimated, and all generators predict a decreasing ratio as a function of multiplicity. However, apart from an initial rise, the measured ratios are approximately constant. For ratios involving Λ baryons, no coherent trend is observed between the two collision systems; for p-He collisions, the measured ratio is slightly decreasing, while it is slightly increasing for p-Ne collisions. In this case, the cross-section ratios are not overestimated by the generators. For p-He collisions, `SIBYLL-2.3d` provides the best description of the data, while `EPOS-LHC` predicts the overall trend in p-Ne collisions best, although it generally underestimates the ratios. As no generator provides a fully consistent description of the data, further improvements to the models are required.

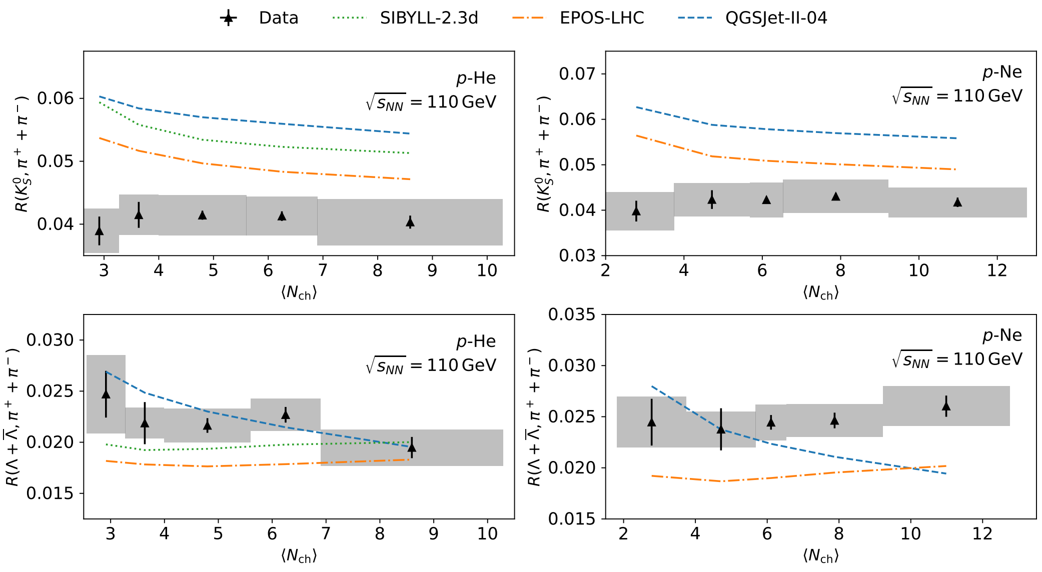


Figure 5.4 – Comparison of the p_{T} integrated measurement with generator predictions. In the top row, the differential cross-section ratios $K^0_{\mathrm{S}}/(\pi^- + \pi^+)$ is presented, while the cross-section ratios $(\Lambda + \overline{\Lambda})/(\pi^- + \pi^+)$ is presented in the bottom row. In both rows, the left column corresponds to p-He collisions, while the right column corresponds to p-Ne collisions.

5.2 Multi-strange particles in SMOG2

Multiplicity-dependent strangeness enhancement has been observed to increase with strangeness content, i.e. the number of valence strange quarks. In particular, findings by the ALICE collaboration show that the cross-section ratios of $\Xi^- + \overline{\Xi}^+$ and $\Omega^- + \overline{\Omega}^+$ to charged pions increase more steeply as a function of multiplicity [31]. Due to limited statistics, a dedicated analysis of multi-strange particles with the available SMOG datasets was not feasible. To enable future measurements using SMOG2, the author has established the groundwork by implementing dedicated HLT2 trigger lines in the LHCb trigger system. The HLT2 lines are essential for the analysis of multi-strange particles as candidates are directly selected and saved during data processing, significantly simplifying future analyses.

The Ξ^- and Ω^- baryons, as well as their antiparticles, are reconstructed by their most dominant decays. For the full decay chain of the Ξ^- baryon, the process $\Xi^- \to \Lambda(\to p\pi^-)\pi^- + cc$ is considered. The overall branching ratio of this process is approximately 64 %, which is obtained by multiplying $\mathcal{B}(\Xi^- \to \Lambda\pi^-) = (99.887 \pm 0.035)\,\%$ [36] and $\mathcal{B}(\Lambda \to p\pi^-) = (64.1 \pm 0.05)\,\%$ [36]. Similarly, the Ω^- baryon is reconstructed using the decay chain $\Omega^- \to \Lambda(\to p\pi^-)K^- + cc$. Given that $\mathcal{B}(\Omega^- \to \Lambda K^-) = (67.7 \pm 0.7)\,\%$ [36], the overall branching ratio is approximately 43.3 %. Leading order Feynman diagrams for the dominant decays of the Ξ^- and Ω^- baryons—not including the decay of the Λ baryon—are presented in Fig. 5.5.

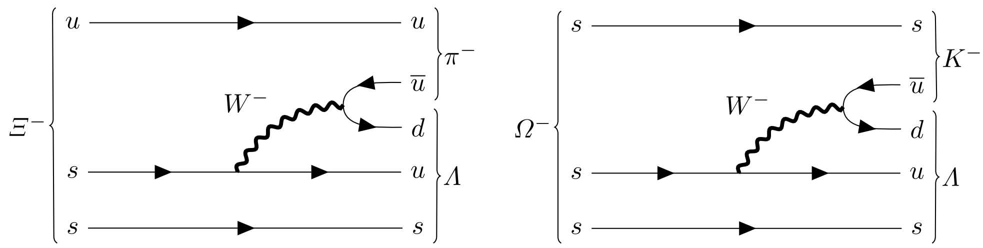


Figure 5.5 – Leading order Feynman diagrams for the decays $\Xi^- \to \Lambda\pi^-$ (left) and $\Omega^- \to \Lambda K^-$(right).

In both cases, the Λ candidate is reconstructed by combining two long tracks originating from the same secondary vertex. A candidate is selected if the combined, invariant mass of the two tracks lies with 25 GeV/c^2 of the known Λ mass. In addition, both tracks are required to satisfy a purely topological selection to ensure that they stem both from the same secondary vertex. Ω^- and Ξ^- candidates are the reconstructed by combining the Λ candidate with an additional long track originating from the same vertex. An additional topological selection is applied to ensure that both the Λ candidate and the track originate from the same vertex. The resulting Ω^- and Ξ^- candidates are required to have an invariant mass within 25 GeV/c^2 of their respective known masses.

In April 2024, the SMOG2 system was commissioned using proton-argon collisions and the HLT2 trigger lines were first applied on data. Argon was specifically chosen as it is the target with the largest atomic number, and thus largest cross-section. The selected candidates, after two hours of data taking, are presented in Fig. 5.6. For Ξ^- candidates, a clear mass peak is visible within the expected region. This is not the case for Ω^- candidates. Here, only what appears to be combinatorial background is visible. This is expected, as the Ω^- baryon has a significantly lower production rate given that it is made up of three valence strange quarks. Using `CHROMO` and `EPOS-LHC`, the production cross-section within LHCb acceptance has been estimated to 0.2656 mb for Ξ^- and 0.0107 mb for Ω^-. This makes Ξ^- production approximately 25 times more likely than Ω^- production, thus explaining the observed differences.

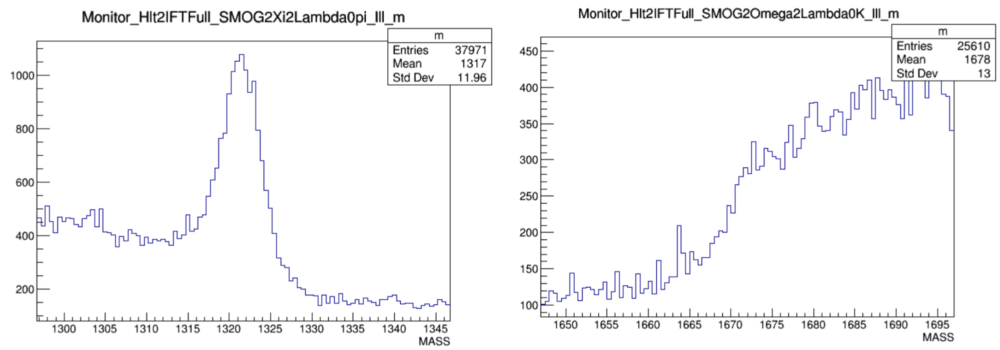


Figure 5.6 – Mass-spectra of multi-strange particles for p-Ar collisions recorded with SMOG2, as selected by the HLT2 trigger lines, from Ref. [127]. Candidates of the decay $\Xi^- \to \Lambda\pi^- + cc$ are shown left and candidates of the decay $\Omega^- \to \Lambda K^- + cc$ are presented right. The candidates are reconstructed from three long-tracks each.

Using LHCb simulation, the HLT2 trigger lines are estimated to reconstruct these multi-strange particles via their outlined decays with an efficiency of 9.2% for Ξ^- and 4.5 % for Ω^- within LHCb acceptance. Since the implementation of Λ reconstruction is equal in both trigger lines, the difference in efficiency likely stems from the different kinematic properties of the π^- and K^- candidates. It is also possible that the different mean lifetimes and masses of the Ξ^- and Ω^- baryons play a role here. However, no further studies nor optimizations have been performed.

5.3 Contributions to analysis preservation

As highlighted in Sec. 2.3.2, the standardized preservation of measurements in machine-readable format within the HEPData and RIVET frameworks is essential for the improvement of hadronic interaction models. In this context, the author contributed by expanding the available analyses and datasets, making them accessible for future studies and model tuning. The preservation of analyses is additionally motivated via the collaborative research centre *SFB 1491* of the German Research Foundations, which the author is associated to. Within this collaborative research centre, multiple projects spanning astrophysics, astroparticle physics, particle physics and plasma physics are integrated with the overarching goal of creating a unifying view on cosmic interacting matter. In two of these subprojects, a focus on hadronic interactions within particle and astroparticle is set, requiring the development of additional RIVET plugins as a step towards global tuning.

As an initial step, the author developed both the HEPData entry and corresponding RIVET plugin for the LHCb measurement of prompt charged particle production in proton-proton collision at $\sqrt{s} = 13$ TeV [128]. The analysis was performed by Dr. habil. Hans Dembinski and Dr. Julian Boelhauve, who were working together with the author in the same working group at TU Dortmund University. During the development, Hans Dembinski consulted the author, resulting in the RIVET plugin published in Ref. [129]. Following this publication, the author was requested by the LHCb simulation group to provide documentation for future plugins, resulting in the guide available in Ref. [130]. This guide was later extended by Julian Boelhauve, who was then a doctoral student at TU Dortmund University, based on the improved understanding of the frameworks gained from the additional projects listed below.

Following the successful development of the RIVET plugin, Julian Boelhauve and the author identified six additional measurements providing relevant information on the hadronic interaction in the atmosphere. It was also ensured that these analyses align with the priorities and needs of the RIVET team. These measurements can be divided into three groups concerning beauty, charm, and charged particle production. The measurements concerning beauty production include the measurement of beauty quark production cross-sections in 7 and 13 TeV proton-proton collisions [131], the beauty hadron production in 8.16 TeV proton-lead collision [132], and the beauty-hadron fractions in 13 TeV proton-proton collisions [133]. The measurements concerning charm production include J/ψ production in proton-lead collisions [134] and prompt D^0 production in proton-proton and proton-lead collisions [135]. The final measurement is regarding the prompt charged particle production in 5 TeV proton-proton and proton-lead collisions [136]. For five out of six measurements, no RIVET plugin existed, while a HEPData entry was missing for three of the six measurements.

The preservation of the above-mentioned measurements was performed in the context of two summer-student projects. Within these projects, Julian Boelhauve and the author supervised Gino Daniels and Joshua Friedman from Purdue University. The summer students prepared first drafts of the RIVET plugins and, where required, HEPData entries. Julian Boelhauve and the author subsequently took care of the review processes until their final publication in Ref. [137–144].

6 Summary and outlook

In this thesis, the production of strange hadrons is studied in proton-helium and proton-neon collisions. The data were collected at the LHCb experiment at a nucleon-nucleon centre-of-mass energy of $\sqrt{s_{\mathrm{NN}}} = 110$ GeV using the SMOG system. Differential cross-section ratios of V^0 particles (K^0_S, $\mathit{\Lambda}$ and $\overline{\mathit{\Lambda}}$) to charged pions ($\pi^+$ and π^-) are measured as functions of centre-of-mass rapidity and transverse momentum. In order to investigate strangeness enhancement as a potential solution to the Muon Puzzle, the analysis is additionally performed in intervals of multiplicity. In doing so, the work presented in this thesis is the first measurement of multiplicity-dependent strangeness production using the SMOG system.

To perform the measurement and enable its extension to multiplicity-dependent observables, a dedicated analysis strategy is developed. A comprehensive procedure is employed to determine intervals of average multiplicity via a mapping of long tracks to promptly produced particles in dedicated calibration and measurement ranges. The cross-section ratios are determined using efficiency-corrected yields of promptly produced V^0 particles and pions. V^0 particles are reconstructed, and their yields determined from invariant mass fits. The V^0 yield is corrected for reconstruction and selection efficiencies. A calibration measurement using 13 TeV proton-proton collisions performed by the LHCb `Tracking Group` is used to account for known differences in data and simulation. This calibration is a leading source of uncertainty, exceeding 8 % in a number of kinematic intervals. Promptly produced charged pions are selected using tracking- and PID information, and their yields are extracted via template fits. They are again corrected for reconstruction and selection efficiencies, with the PID efficiency determined using data-driven methods. This constitutes the second-largest source of uncertainty, exceeding the uncertainty from the calibration procedure in certain kinematic regions. For all particles, backgrounds have been estimated using the simulated samples. The resulting double-differential cross-section ratios are compared to three state-of-the-art hadronic event generators, `EPOS-LHC`, `SIBYLL-2.3d`, and `QGSJet-II-04`, none of which were able to consistently reproduce the measured observables within its uncertainties.

For the comparison with the findings of the ALICE collaboration, the multiplicity-dependent measurement is integrated with respect to transverse momentum, resulting in a single-differential observable in average multiplicity. Significant, multiplicity dependent strangeness production is observed for neither K^0_S mesons nor $\mathit{\Lambda}$ and $\overline{\mathit{\Lambda}}$ baryons. For the Muon Puzzle, this implies that strangeness enhancement at a nucleon-nucleon centre-of-mass energy of 110 GeV is unlikely to account for the observed muon excess in air showers.

Once this measurement is published, it can be preserved in both the HEPData and RIVET frameworks, complementing the author's previously published entries and plugins. Making the results publicly available within these frameworks would enhance both visibility and impact of this measurement. This measurement could then be extended to other SMOG datasets, thus providing the observables for different beam-energies and targets. Adapting this analysis to the newly obtained SMOG2 samples would be of particular interest. With the increased available statistics, the calibration measurement could be directly performed on SMOG data, thus potentially reducing the overall uncertainties. In addition, luminosity can be precisely determined for the SMOG2 samples, enabling the measurement of total differential cross-sections. Lastly, the HLT2 lines developed by the author could be employed to study the multiplicity dependent production of multi-strange particles in SMOG2, which are expected to rise more steeply as a function of multiplicity.

Although no significant strangeness enhancement is observed, the measurement provides important inputs for the tuning of event generators, which are particularly relevant for the modelling of cosmic-ray propagation in Earth's atmosphere. The outlined future opportunities promise to further enhance both the precision and impact of similar measurements, providing important inputs for hadronic interaction models and a deeper understanding of strangeness production across different energies and collision systems.

A Appendix

A.1 Definition of long-lived particles

Table A.1 – List of long-lived particles ($\tau > 30\,\mathrm{ps}$) as described in Ref [145].

particle	charge	$\tau/\,\mathrm{ps}$
e	-1	∞
μ	-1	$2.197 \cdot 10^{6}$
gluon	0	∞
γ	0	∞
K_{L}^{0}	0	51143
π^{+}	+1	26033
K_{S}^{0}	0	89.5
K^{+}	+1	12379
n	0	$8.794 \cdot 10^{14}$
p	+1	∞
Σ^{-}	-1	147.9
Λ	0	263.2
Σ^{+}	+1	80.2
Ξ^{-}	-1	163.7
Ξ^{0}	0	290.0
Ω^{-}	-1	82.1

A.2 Software

Apart from the ROOT [146] and LHCb [147, 148] software frameworks used in the initial steps in the analysis, the corresponding code to this analysis is entirely written in the PYTHON programming language. The following packages, including standard scientific packages and high-energy-physics-specific packages from the SCIKIT-HEP project [149], are used:

- NUMPY [150]: general-purpose array-processing package
- PANDAS [151, 152]: data structures
- MATPLOTLIB [153]: plotting library
- CHROMO [125]: interface to hadronic-interaction models
- JACOBI [154]: fast numerical derivatives and propagation of uncertainty
- NUMBA [155]: just-in-time compiler for PYTHON code
- NUMBA-STATS [156]: NUMBA-accelerated implementations of statistical distributions
- SCIKIT-LEARN [157]: machine-learning tools
- SCIPY [158]: statistical distributions and interpolation routines
- SNAKEMAKE [159]: workflow engine
- SYMPY [160]: computer algebra system
- XGBOOST [120]: optimised gradient-boosted decision trees
- AWKWARD ARRAY [161]: operations on nested variable-length arrays
- BOOST-HISTOGRAM [162]: multidimensional histograms and profiles
- IMINUIT [163]: fitting statistical models
- PARTICLE [164]: Particle Data Group tables and Monte Carlo particle numbers
- UPROOT [165]: reader and writer of ROOT file format in pure PYTHON
- VECTOR [166] JIT-compilable mathematical manipulations of ragged Lorentz vectors
- POETRY [167]: virtual environment management

A.3 Description of two body decays

The two-body V^0 decay into two charged particles can be fully described using the position of the primary vertex $\vec{PV}$, the positions $\vec{r}_1$ and $\vec{r}_2$ of the charged particle trajectories, and the momentum vectors $\vec{p}_1$ and $\vec{p}_2$ of the charged particles at $\vec{r}_1$ and $\vec{r}_2$. Due to translation invariance, the coordinate system can be chosen such as $\vec{PV} = \vec{0}$. Using the assumption that tracks are straight lines and that the momentum does not change, the vectors $\vec{r}$ can be replaced by the vectors $\vec{q}$, which are orthogonal to $\vec{p}$:

$$\vec{q} = \vec{r} - \frac{\vec{r} \cdot \vec{p}}{p^2} \vec{p}. \tag{A.1}$$

By also assuming rotation symmetry, the number of independent variables can be further reduced. This is done by computing rotation invariant quantities, which are, in 3D space, scalars and pseudo-scalars. The full set of variables is given by all pairs and triplets from $\{\vec{p}_1, \vec{p}_2, \vec{q}_1, \vec{q}_2\}$:

$$p_{ij} = \vec{p}_i \cdot \vec{p}_j \tag{A.2}$$

$$Q_{ij} = \vec{q}_i \cdot \vec{p}_j \tag{A.3}$$

$$q_{ij} = \vec{q}_i \cdot \vec{q}_j \tag{A.4}$$

$$A_i = \vec{q}_i \cdot (\vec{p}_1 \times \vec{p}_2) \tag{A.5}$$

$$B_i = \vec{p}_i \cdot (\vec{q}_1 \times \vec{q}_2), \tag{A.6}$$

with $i, j \in \{1, 2\}$. Out of these, only the following seven are independent: p_{11}, p_{12}, p_{22}, Q_{12}, Q_{21}, A_1, and A_2. These variables can then be used to construct six geometric variables, which can be interpreted as physical quantities. To simplify the notation of these variables, the term $N^2 = p_{11}p_{22} - p_{12}^2$ is defined. The first three variables are the distance of closest approach between the V^0 decay products, D, the distance of the primary vertex to the event plane, S, and the in-plane distance impact-parameter of the V^0 to the primary vertex, I:

$$D = \frac{\mid A_1 - A_2 \mid}{N} \tag{A.7}$$

$$S = \frac{\mid A_1 + A_2 \mid}{N} \tag{A.8}$$

$$I = \frac{\mid p_{22}Q_{21} - p_{11}Q_{12} \mid}{N\sqrt{p_{11} + p_{22} + 2p_{12}}}, \tag{A.9}$$

which are ideally close to zero. The remaining variables are the in-plane squared impact parameters of the tracks from the primary vertex, I_{11} and I_{22}, and a variable containing the information to the flight-distance not covered by the other variables,

I_{12}, which are all sensitive to non-zero lifetimes:

$$I_{11} = \frac{p_{11}Q_{12}^2}{N^2} \tag{A.10}$$

$$I_{22} = \frac{p_{22}Q_{21}^2}{N^2} \tag{A.11}$$

$$I_{12} = \frac{p_{12}Q_{21}Q_{12}}{N^2}. \tag{A.12}$$

The seventh and final independent variable would be the invariant mass of the V^0 candidate. However, this variable is excluded from the BDT training to avoid trivial discrimination based solely on mass.

A.4 BDT input variables

The geometric input variables are shown for Λ and $\overline{\Lambda}$ particles for the p-He simulated sample.

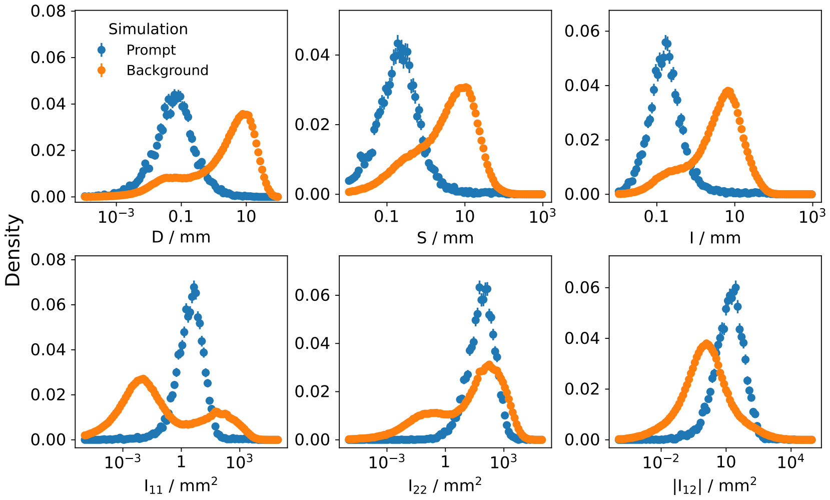


Figure A.1 – Distribution of BDT input variables for $\Lambda \to p\pi^-$ candidates in the p-He simulation sample. Here, signal refers to prompt Λ particles, whereas background refers to everything else.

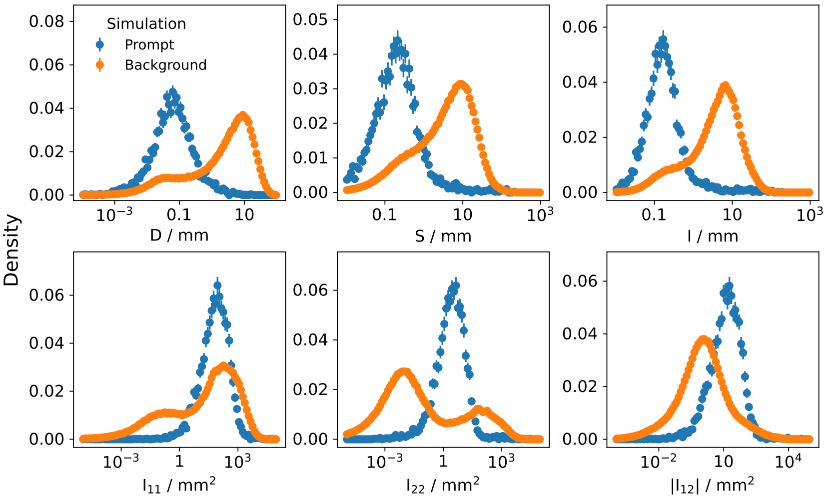


Figure A.2 – Distribution of BDT input variables for $\overline{\Lambda} \to \overline{p}\pi^+$ candidates in the p-He simulation sample. Here, signal refers to prompt $\overline{\Lambda}$ particles, whereas background refers to everything else.

A.5 Tracking efficiency correction for the 2017 samples

In this section, the tracking efficiency correction corresponding to the p-Ne 2017 samples are presented.

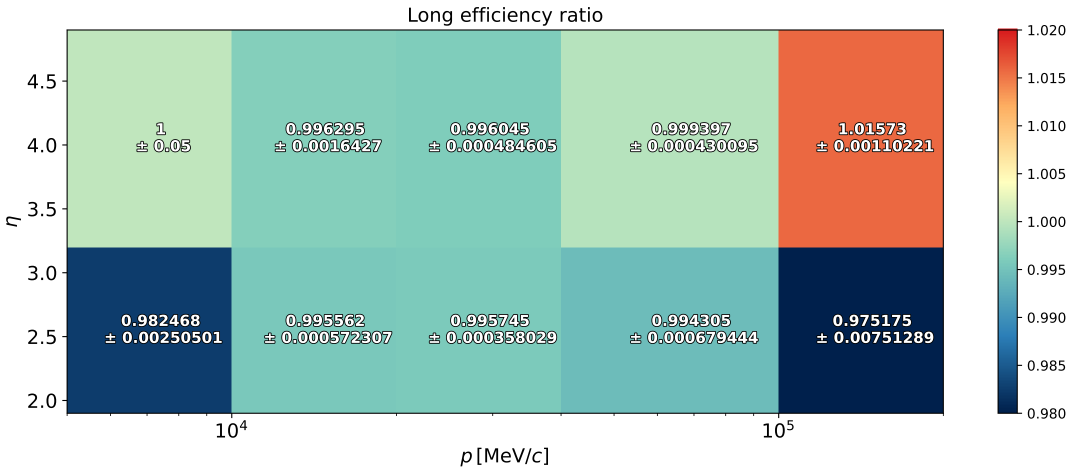


Figure A.3 – Map of correction factors to account for discrepancies between real and simulated samples for 2017 data-taking conditions. The values are provided by the `Tracking Group` [123], using the procedure described in Ref. [118].

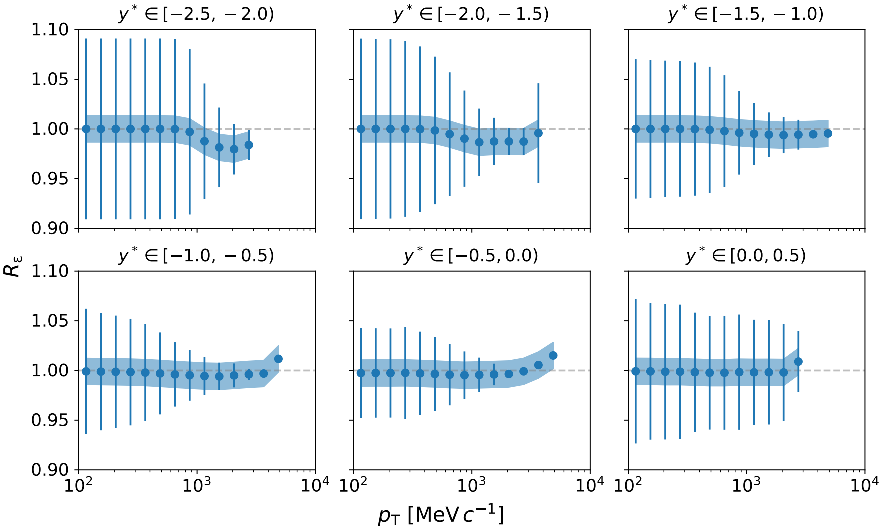


Figure A.4 – Tracking correction for $K^0_S \rightarrow \pi^+\pi^-$ decays for the p-Ne simulated sample. The bars and boxes represent the statistical and systematic uncertainty, respectively. The grey dashed line serves as a visual guide, indicating the value of 1.

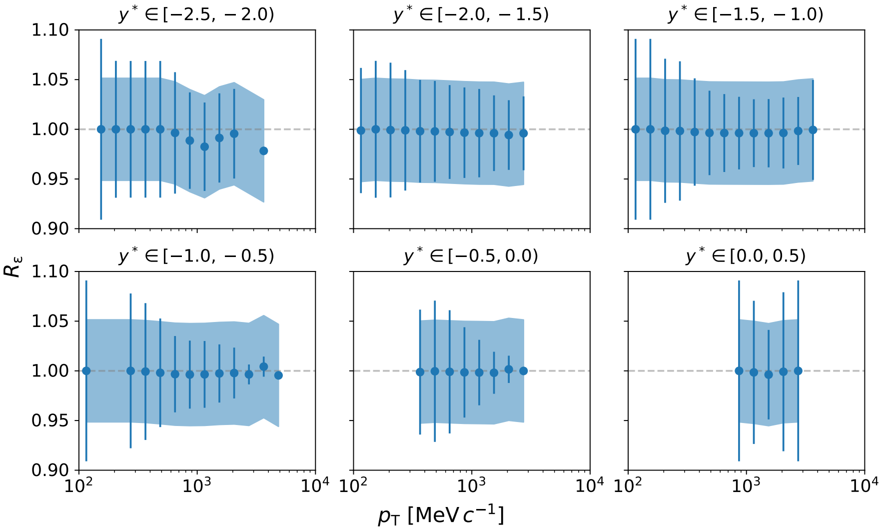


Figure A.5 – Tracking correction for $\Lambda \rightarrow p\pi^-$ decays for the p-Ne simulated sample. The bars and boxes represent the statistical and systematic uncertainty, respectively. The grey dashed line serves as a visual guide, indicating the value of 1.

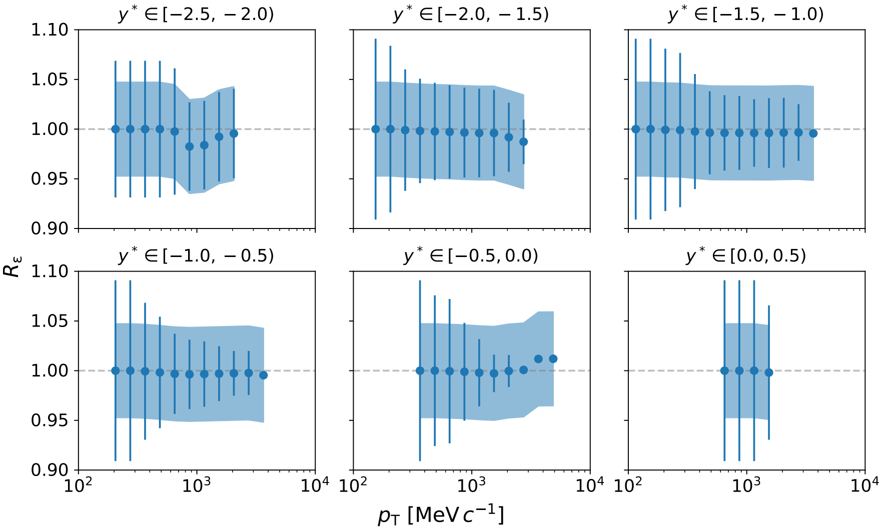


Figure A.6 – Tracking correction for $\overline{\Lambda} \to \overline{p}\pi^+$ decays for the p-Ne simulated sample. The bars and boxes represent the statistical and systematic uncertainty, respectively. The grey dashed line serves as a visual guide, indicating the value of 1.

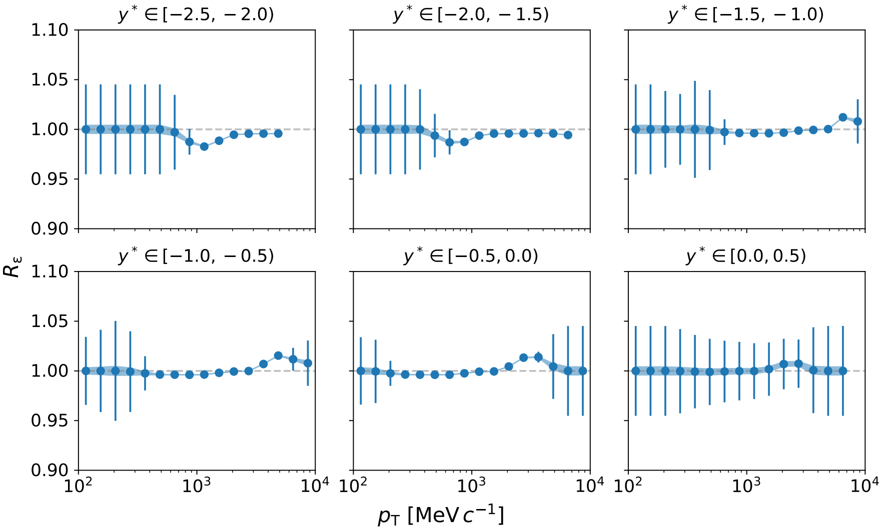


Figure A.7 – Tracking correction for prompt π^- particles for the p-Ne simulated sample. The bars and boxes represent the statistical and systematic uncertainty, respectively. The grey dashed line serves as a visual guide, indicating the value of 1.

A.6 Results obtained from the proton-neon samples

A.6.1 Cross-section ratios of K_S^0 to charged pions

In this section, the differential cross-section ratios $K_\mathrm{S}^0/(\pi^+ + \pi^-)$ obtained from the p-Ne dataset are presented as a function of y^* and p_T, and are illustrated in Fig. A.8. As observed previously in Sec. 4.6.1, the measurement is systematically dominated. For most intervals, the systematic uncertainties range from 4 to 12 %, while the statistical uncertainties remain below 3 %. Overall, the generators tend to overestimate the ratio across most kinematic intervals, while the general trends are best reproduced by `EPOS-LHC`. The predictions by `EPOS-LHC` and `QGSJet-II-04` mimic those discussed in Sec. 4.6.1. The ratio as predicted by `EPOS-LHC` starts lower than that from `QGSJet-II-04` and rises more steeply until around 1 GeV/c, after which the slope flattens. This aligns well with the trends observed in data. In contrast, `QGSJet-II-04` start off higher at low p_T and continuously increase, with no noticeable reduction in slope towards higher p_T. Again, there exist a few intervals in which predictions by either `EPOS-LHC` or `QGSJet-II-04` are compatible with the measurement within uncertainties.

The ratios as a function of p_T and $\langle N_\mathrm{ch} \rangle$ are presented in Fig. A.9. Again, similar observations to those reported in Sec. 4.6.1 are made. For most intervals, the systematic uncertainties range from 5 to 15 %, while the statistical uncertainties remain below 3 %. At both low and high p_T, the overall trends are described well by the generators. At low p_T, both `EPOS-LHC` and `QGSJet-II-04` overestimate the ratio. For high p_T however, the predictions of `EPOS-LHC` are within the uncertainties of the measurement, while `QGSJet-II-04` overestimates the ratio again. At mid p_T, the overall trends as observed by the generator predictions do not agree with the measurement; again, the measured ratio rises as a function of multiplicity, while both generators predict a slightly decreasing ratio. Towards high multiplicity, both generator predictions are within the uncertainties of the measurement, while they overestimate it for low multiplicity.

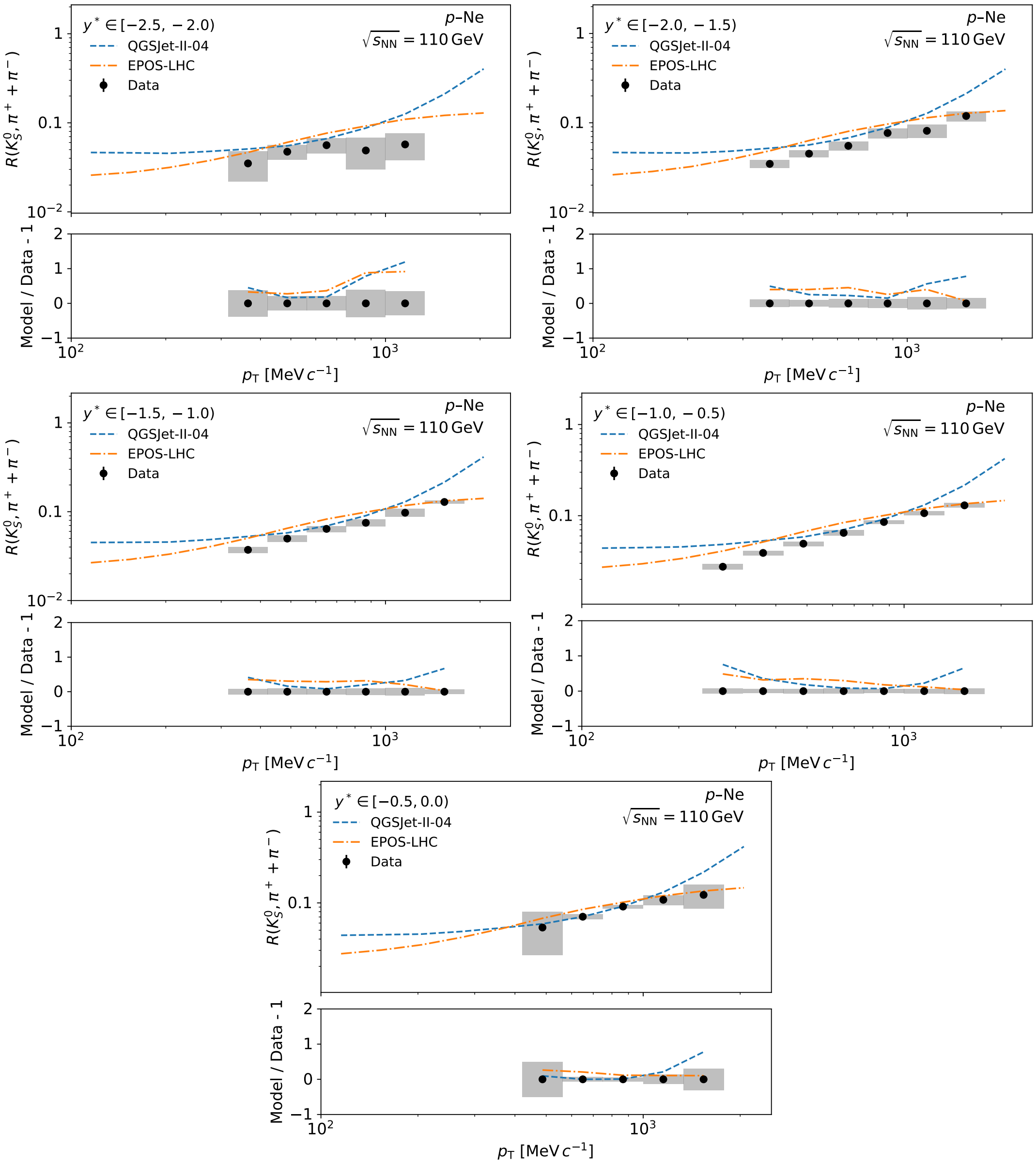


Figure A.8 – Differential cross-section ratios of $K^0_{\rm S}$ particles to charged pions, π^- and π^+, in intervals of y^* and $p_{\rm T}$ for the p-Ne 2017 dataset. The black points show the measured ratios, and coloured lines correspond to the predictions of different event generators. The bars and boxes represent the statistical and systematic uncertainty, respectively. The relative deviation of the models to the measured ratio is given in the lower part of each plot.

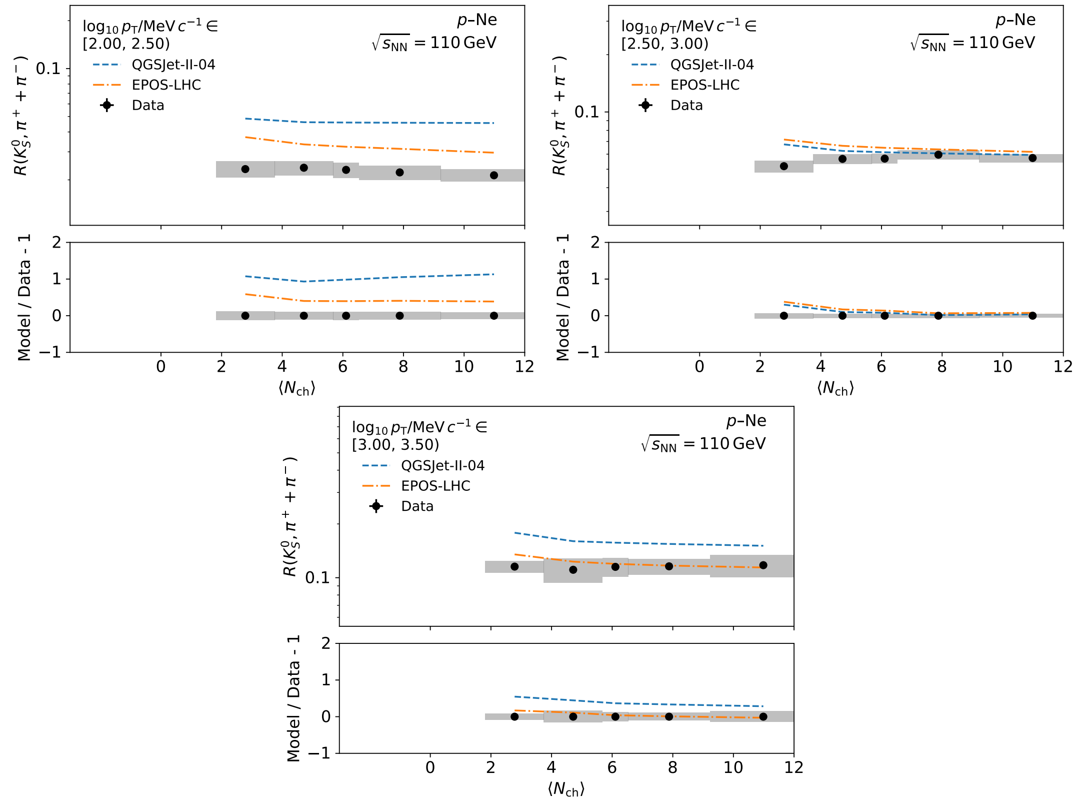


Figure A.9 – Differential cross-section ratios of K_S^0 particles to charged pions, π^- and π^+, in intervals of p_T and $\langle N_{ch} \rangle$ for the p-Ne 2017 dataset. The black points show the measured ratios, and coloured lines correspond to the predictions of different event generators. The bars and boxes represent the statistical and systematic uncertainty, respectively. The relative deviation of the models to the measured ratio is given in the lower part of each plot.

A.6.2 Cross-section ratios of Λ and $\overline{\Lambda}$ to charged pions

In this section, the differential cross-section ratios $(\Lambda+\overline{\Lambda})/(\pi^{+}+\pi^{-})$ derived from the p-Ne dataset are presented. The ratios in intervals in y^{*} and p_{T} are presented in Fig. A.10. For this ratio, systematic uncertainties dominate the measurement. For most intervals, the systematic uncertainties range from 5 to 13 %, with the uncertainty increasing towards the kinematic edges. The statistical uncertainties range between 5 to 12 % for the majority of intervals, again increasing towards the kinematic edges. Overall, the ratios are increasing as a function of transverse momentum, a trend predicted by both `EPOS-LHC` and `QGSJet-II-04`. Again, the predictions mainly differ in the predicted slope towards high p_{T}; the ratio predicted by `QGSJet-II-04` exhibits a steeper slope towards higher p_{T}, while the slope predicted by `EPOS-LHC` is decreasing. The slope is described best by `QGSJet-II-04`, although the predictions tend to slightly overestimate the ratio.

The differential cross-section ratio in intervals of p_{T} and $\langle N_{\mathrm{ch}}\rangle$ is presented in Fig. A.11. Again, the ratios are systematically dominated. For most intervals, the systematic uncertainties range from 5 to 17 %, with the uncertainty increasing towards the kinematic edges. The statistical uncertainties range between 2 to 12 % for the majority of intervals, again increasing towards the kinematic edges. For low p_{T}, the ratio is decreasing as a function of multiplicity which is well reproduced by both generators. At mid p_{T}, the measured ratio is slightly rising as a function of multiplicity, followed by a subsequent decrease. However, due to the large uncertainties, this trend is not significant. The generators predict either a slightly increasing or decreasing ratio, but underestimate it overall. At high p_{T}, the measured ratio is again slightly increasing as a function of multiplicity. Again, this increase is not significant due to the large uncertainties. Although `QGSJet-II-04` predicts a slightly decreasing ratio, it remains compatible with the measurement within the uncertainties, while `EPOS-LHC` underestimates the ratio.

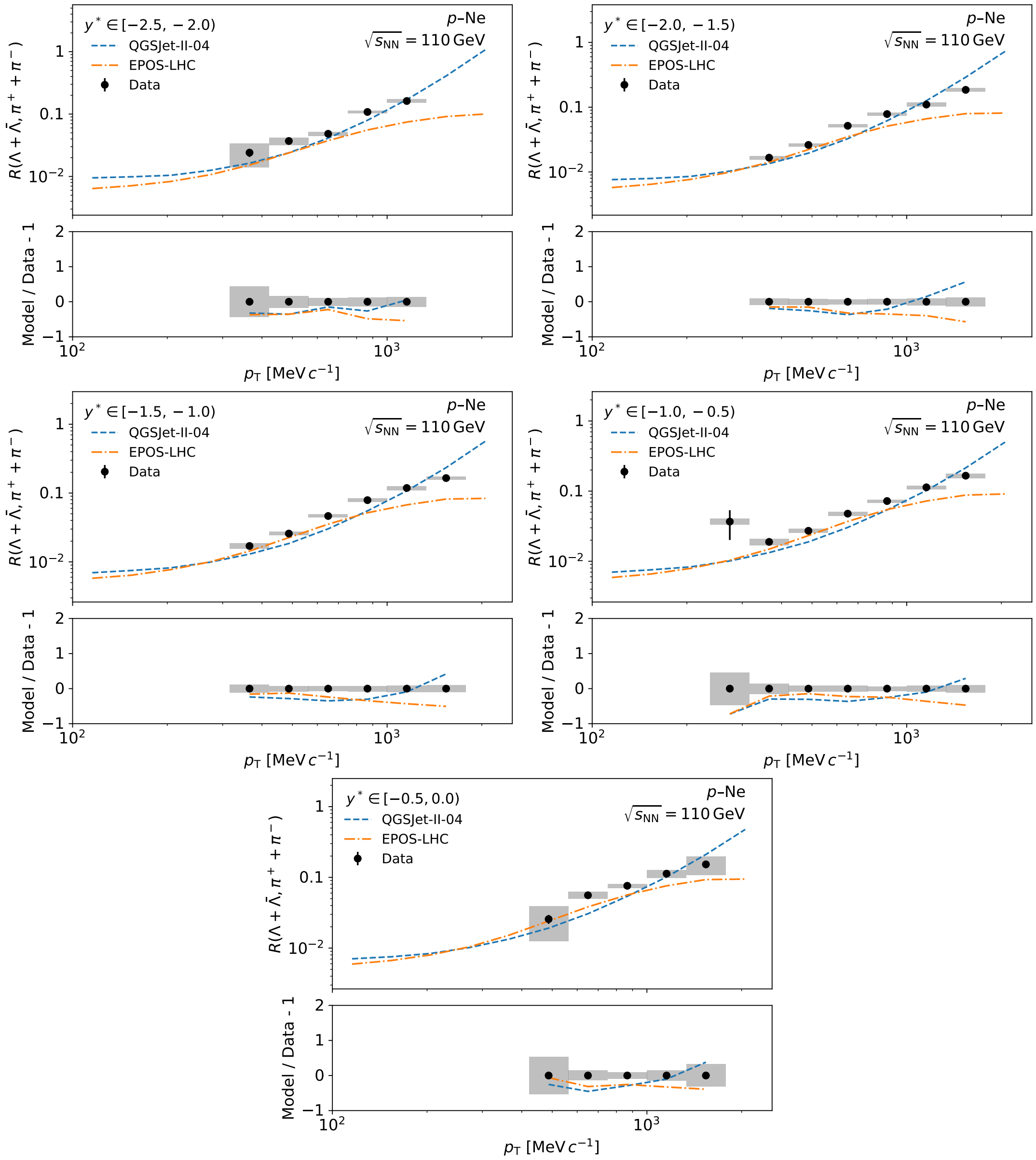


Figure A.10 – Differential cross-section ratios of Λ and $\overline{\Lambda}$ to charged pions, π^- and π^+, in intervals of y^* and p_{T} for the p-Ne 2017 dataset. The black points show the measured ratios, and coloured lines correspond to the predictions of different event generators. The bars and boxes represent the statistical and systematic uncertainty, respectively. The relative deviation of the models to the measured ratio is given in the lower part of each plot.

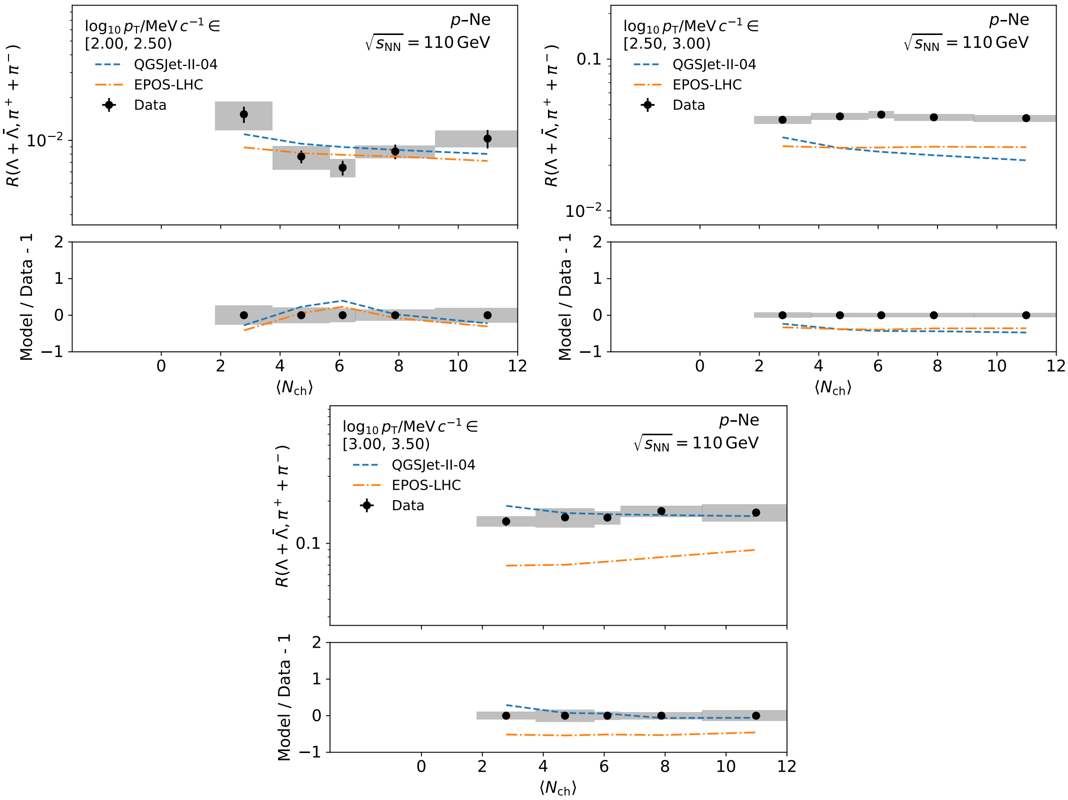


Figure A.11 – Differential cross-section ratios of Λ and $\overline{\Lambda}$ particles to charged pions, π^- and π^+, in intervals of p_T and $\langle N_{ch} \rangle$ for the p-Ne 2017 dataset. The black points show the measured ratios, and coloured lines correspond to the predictions of different event generators. The bars and boxes represent the statistical and systematic uncertainty, respectively. The relative deviation of the models to the measured ratio is given in the lower part of each plot.

Bibliography

[1] T. K. Gaisser, R. Engel and E. Resconi, *Cosmic Rays and Particle Physics*, 2nd ed., Cambridge University Press, 2016.

[2] D. Skobelzyn, *Über eine neue Art sehr schneller β-Strahlen*, Zeitschrift für Physik **54**.9 (1929) 686–702, DOI: `10.1007/BF01341600`.

[3] C. D. Anderson, *The Apparent Existence of Easily Deflectable Positives*, Science **76** (1932) 238–239, DOI: `10.1126/science.76.1967.238`.

[4] C. D. Anderson, *The Positive Electron*, Phys. Rev. **43** (1933) 491–494, DOI: `10.1103/PhysRev.43.491`.

[5] Nobel Prize Outreach 2025, *Nobel Prize in Physics 1936*, 2025, `https://www.nobelprize.org/prizes/physics/1936/summary/` (visited on 01/10/2025).

[6] S. H. Neddermeyer and C. D. Anderson, *Note on the Nature of Cosmic-Ray Particles*, Phys. Rev. **51** (10 1937) 884–886, DOI: `10.1103/PhysRev.51.884`.

[7] G. D. Rochester and C. C. Butler, *Evidence for the Existence of New Unstable Elementary Particles*, Nature **160** (1947) 855–857, DOI: `10.1038/160855a0`.

[8] R. Armenteros *et al.*, *Decay of V-Particles*, Nature **167**.4248 (1951) 501–503, DOI: `10.1038/167501a0`.

[9] V. D. Hopper and S. Biswas, *Evidence Concerning the Existence of the New Unstable Elementary Neutral Particle*, Phys. Rev. **80** (6 1950) 1099–1100, DOI: `10.1103/PhysRev.80.1099`.

[10] A. J. Seriff *et al.*, *Cloud-Chamber Observations of the New Unstable Cosmic-Ray Particles*, Phys. Rev. **78** (3 1950) 290–291, DOI: `10.1103/PhysRev.78.290`.

[11] C. M. G. Lattes *et al.*, *PROCESSES INVOLVING CHARGED MESONS*, Nature **159** (1947) 694–697, DOI: `10.1038/159694a0`.

[12] D. H. Perkins, *Nuclear Disintegration by Meson Capture*, Nature **159**.4030 (1947) 126–127, DOI: `10.1038/159126a0`.

[13] M. H. Blewett, *The Cosmotron—A Review*, Rev. Sci. Instrum. **24**.9 (1953) 725–737, DOI: `10.1063/1.1770809`.

[14] A. Pais, *Some Remarks on the V-Particles*, Phys. Rev. **86** (5 1952) 663–672, DOI: `10.1103/PhysRev.86.663`.

[15] A. Pais, *On the Baryon-meson-photon System*, Progress of Theoretical Physics **10**.4 (1953) 457–469, ISSN: 0033-068X, DOI: `10.1143/PTP.10.457`.

[16] M. Gell-Mann, *Isotopic Spin and New Unstable Particles*, Phys. Rev. **92** (3 1953) 833–834, DOI: `10.1103/PhysRev.92.833`.

[17] M. Gell-Mann, *The interpretation of the new particles as displaced charge multiplets*, Il Nuovo Cimento (1955-1965) **4**.2 (1956) 848–866, DOI: `10.1007/BF02748000`.

[18] R. Bjorklund *et al.*, *High Energy Photons from Proton-Nucleon Collisions*, Phys. Rev. **77** (1950) 213–218, DOI: `10.1103/PhysRev.77.213`.

[19] O. Chamberlain *et al.*, *Observation of Anti-protons*, Phys. Rev. **100** (1955) 947–950, DOI: `10.1103/PhysRev.100.947`.

[20] G. Danby *et al.*, *Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos*, Phys. Rev. Lett. **9** (1962) 36–44, DOI: `10.1103/PhysRevLett.9.36`.

[21] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579–588, DOI: `10.1016/0029-5582(61)90469-2`.

[22] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964), ed. by J. C. Taylor 321–323, DOI: `10.1103/PhysRevLett.13.321`.

[23] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266, DOI: `10.1103/PhysRevLett.19.1264`.

[24] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519** (1968) 367–377, DOI: `10.1142/9789812795915_0034`.

[25] G. 't Hooft and M. J. G. Veltman, *Regularization and Renormalization of Gauge Fields*, Nucl. Phys. B **44** (1972) 189–213, DOI: `10.1016/0550-3213(72)90279-9`.

[26] F. Halzen and A. D. Martin, *QUARKS AND LEPTONS: An Introductory Course In Modern Particle Physics*, 1984, ISBN: 978-0-471-88741-6.

[27] S. Weinberg, *Essay: Half a Century of the Standard Model*, Phys. Rev. Lett. **121** (22 2018) 220001, DOI: `10.1103/PhysRevLett.121.220001`.

[28] G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1–29, DOI: `10.1016/j.physletb.2012.08.020`.

[29] S. Chatrchyan *et al.*, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716** (2012) 30–61, DOI: `10.1016/j.physletb.2012.08.021`.

[30] J. Albrecht *et al.*, *The Muon Puzzle in cosmic-ray induced air showers and its connection to the Large Hadron Collider*, Astrophys. Space Sci. **367**.3 (2022) 27, DOI: `10.1007/s10509-022-04054-5`.

[31] J. Adam *et al.*, *Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions*, Nature Phys. **13** (2017) 535–539, DOI: `10.1038/nphys4111`.

[32] B. I. Abelev *et al.*, *Enhanced strange baryon production in Au + Au collisions compared to p + p at* $\sqrt{s_{NN}} = 200\,GeV$, Phys. Rev. C **77** (2008) 044908, DOI: `10.1103/PhysRevC.77.044908`.

[33] F. Antinori *et al.*, *Energy dependence of hyperon production in nucleus nucleus collisions at SPS*, Phys. Lett. B **595** (2004) 68–74, DOI: `10.1016/j.physletb.2004.05.025`.

[34] Wikimedia Commons, *Standard Model of Elementary Particles*, Wikimedia Commons — uploaded 17 September 2019., 2019, `https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg` (visited on 22/10/2025).

[35] I. Neutelings, *Standard Model*, 2024, `https://tikz.net/sm_particles/` (visited on 22/10/2025).

[36] S. Navas *et al.*, *Review of particle physics*, Phys. Rev. D **110**.3 (2024) 030001, DOI: `10.1103/PhysRevD.110.030001`.

[37] Y. Ashie *et al.*, *Evidence for an oscillatory signature in atmospheric neutrino oscillation*, Phys. Rev. Lett. **93** (2004) 101801, DOI: `10.1103/PhysRevLett.93.101801`.

[38] M. Gell-Mann, *Symmetries of baryons and mesons*, Phys. Rev. **125** (1962) 1067–1084, DOI: `10.1103/PhysRev.125.1067`.

[39] R. Aaij *et al.*, *Observation of a narrow pentaquark state,* $P_c(4312)^+$*, and of two-peak structure of the* $P_c(4450)^+$, Phys. Rev. Lett. **122**.22 (2019) 222001, DOI: `10.1103/PhysRevLett.122.222001`.

[40] R. Aaij *et al.*, *Observation of a* $J/\psi\Lambda$ *Resonance Consistent with a Strange Pentaquark Candidate in* $B\text{-}{\longrightarrow}J/\psi\Lambda\bar{p}$ *Decays*, Phys. Rev. Lett. **131**.3 (2023) 031901, DOI: `10.1103/PhysRevLett.131.031901`.

[41] R. Aaij *et al.*, *Observation of the Open-Charm Tetraquark Candidate* $T^{*}_{cs0}(2870)^0$ *in the* $B^- \to D^- D^0 K^0_S$ *Decay*, Phys. Rev. Lett. **134**.10 (2025) 101901, DOI: `10.1103/PhysRevLett.134.101901`.

[42] F. Barao, *AMS: Alpha Magnetic Spectrometer on the International Space Station*, Nucl. Instrum. Meth. A **535** (2004), ed. by M. Jeitler *et al.* 134–138, DOI: `10.1016/j.nima.2004.07.196`.

[43] M. Bouffard *et al.*, *The HEAO-3 Cosmic Ray Isotope spectrometer*, Astrophysics and Space Science **84**.1 (1982) 3–33, DOI: `10.1007/BF00713625`.

[44] E. Seo *et al.*, *Cosmic-ray energetics and mass (CREAM) balloon project*, Advances in Space Research **33**.10 (2004), The Next Generation in Scientific Ballooning 1777–1785, ISSN: 0273-1177, DOI: `10.1016/j.asr.2003.05.019`.

[45] A. Aab *et al.*, *The Pierre Auger Cosmic Ray Observatory*, Nucl. Instrum. Meth. A **798** (2015) 172–213, DOI: `10.1016/j.nima.2015.06.058`.

[46] T. Abu-Zayyad *et al.*, *The surface detector array of the Telescope Array experiment*, Nucl. Instrum. Meth. A **689** (2013) 87–97, DOI: `10.1016/j.nima.2012.05.079`.

[47] K.-H. Kampert, A. A. Watson and A. A. Watson, *Extensive Air Showers and Ultra High-Energy Cosmic Rays: A Historical Review*, Eur. Phys. J. H **37** (2012) 359–412, DOI: `10.1140/epjh/e2012-30013-x`.

[48] P. Auger *et al.*, *Extensive cosmic ray showers*, Rev. Mod. Phys. **11** (1939) 288–291, DOI: `10.1103/RevModPhys.11.288`.

[49] A. Haungs *et al.*, *KCDC - The KASCADE Cosmic-ray Data Centre*, J. Phys. Conf. Ser. **632**.1 (2015) 012011, DOI: `10.1088/1742-6596/632/1/012011`.

[50] *Auger Open Data — Event Browser*, `https://opendata.auger.org/display.php`, Accessed: 2025-11-01. Release version 3, Mar 20 2024, Pierre Auger Collaboration, 2024.

[51] V. N. Gribov, *A Reggeon Diagram Technique*, Sov. Phys. JETP **26** (1968) 414–423.

[52] V. N. Gribov, *Glauber Corrections and the Interaction between High-energy Hadrons and Nuclei*, Sov. Phys. JETP **29** (1969) 483–487.

[53] H. J. Drescher *et al.*, *Parton based Gribov-Regge theory*, Phys. Rept. **350** (2001) 93–289, DOI: `10.1016/S0370-1573(00)00122-8`.

[54] M. Kachelrieß, I. V. Moskalenko and S. Ostapchenko, *AAfrag: Interpolation routines for Monte Carlo results on secondary production in proton-proton, proton-nucleus and nucleus-nucleus interactions*, Comput. Phys. Commun. **245** (2019) 106846, DOI: `10.1016/j.cpc.2019.08.001`.

[55] T. Pierog *et al.*, *EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider*, Phys. Rev. C **92**.3 (2015) 034906, DOI: `10.1103/PhysRevC.92.034906`.

[56] F. Riehn *et al.*, *Hadronic interaction model Sibyll 2.3d and extensive air showers*, Phys. Rev. D **102**.6 (2020) 063002, DOI: `10.1103/PhysRevD.102.063002`.

[57] T. Sjostrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852–867, DOI: `10.1016/j.cpc.2008.01.036`.

[58] B. Andersson *et al.*, *Parton Fragmentation and String Dynamics*, Phys. Rept. **97** (1983) 31–145, DOI: `10.1016/0370-1573(83)90080-7`.

[59] A. Ortiz Velasquez *et al.*, *Color Reconnection and Flowlike Patterns in pp Collisions*, Phys. Rev. Lett. **111**.4 (2013) 042001, DOI: `10.1103/PhysRevLett.111.042001`.

[60] C. Bierlich, G. Gustafson and L. Lönnblad, *Collectivity without plasma in hadronic collisions*, Phys. Lett. B **779** (2018) 58–63, DOI: `10.1016/j.physletb.2018.01.069`.

[61] C. Bierlich *et al.*, *Setting the string shoving picture in a new frame*, JHEP **03** (2021) 270, DOI: `10.1007/JHEP03(2021)270`.

[62] C. Bierlich *et al.*, *Strangeness enhancement across collision systems without a plasma*, Phys. Lett. B **835** (2022) 137571, DOI: `10.1016/j.physletb.2022.137571`.

[63] C. Bierlich *et al.*, *Effects of Overlapping Strings in pp Collisions*, JHEP **03** (2015) 148, DOI: `10.1007/JHEP03(2015)148`.

[64] K. Werner, *Core-corona separation in ultra-relativistic heavy ion collisions*, Phys. Rev. Lett. **98** (2007) 152301, DOI: `10.1103/PhysRevLett.98.152301`.

[65] K. Werner, F.-M. Liu and T. Pierog, *Parton ladder splitting and the rapidity dependence of transverse momentum spectra in deuteron-gold collisions at RHIC*, Phys. Rev. C **74** (2006) 044902, DOI: `10.1103/PhysRevC.74.044902`.

[66] A. Capella *et al.*, *Dual parton model*, Phys. Rept. **236** (1994) 225–329, DOI: `10.1016/0370-1573(94)90064-7`.

[67] J. Albrecht *et al.*, *Global tuning of hadronic interaction models with accelerator-based and astroparticle data*, Nature Reviews Physics (2025), DOI: `10.1038/s42254-025-00897-3`.

[68] R. Engel, D. Heck and T. Pierog, *Extensive air showers and hadronic interactions at high energy*, Ann. Rev. Nucl. Part. Sci. **61** (2011) 467–489, DOI: `10.1146/annurev.nucl.012809.104544`.

[69] T. Pierog, *Air Shower Simulation with a New Generation of post-LHC Hadronic Interaction Models in CORSIKA*, PoS **ICRC2017** (2018) 1100, DOI: `10.22323/1.301.1100`.

[70] A. Aab *et al.*, *Muons in Air Showers at the Pierre Auger Observatory: Mean Number in Highly Inclined Events*, Phys. Rev. D **91**.3 (2015), [Erratum: Phys.Rev.D 91, 059901 (2015)] 032003, DOI: `10.1103/PhysRevD.91.032003`.

[71] A. Aab *et al.*, *Testing Hadronic Interactions at Ultrahigh Energies with Air Showers Measured by the Pierre Auger Observatory*, Phys. Rev. Lett. **117**.19 (2016) 192001, DOI: `10.1103/PhysRevLett.117.192001`.

[72] H. P. Dembinski *et al.*, *Report on Tests and Measurements of Hadronic Interaction Properties with Air Showers*, EPJ Web Conf. **210** (2019), ed. by I. Lhenry-Yvon *et al.* 02004, DOI: `10.1051/epjconf/201921002004`.

[73] A. Aab *et al.*, *Measurement of the Fluctuations in the Number of Muons in Extensive Air Showers with the Pierre Auger Observatory*, Phys. Rev. Lett. **126**.15 (2021) 152002, DOI: `10.1103/PhysRevLett.126.152002`.

[74] S. Baur *et al.*, *Core-corona effect in hadron collisions and muon production in air showers*, Phys. Rev. D **107**.9 (2023) 094031, DOI: `10.1103/PhysRevD.107.094031`.

[75] C. Bierlich and J. R. Christiansen, *Effects of color reconnection on hadron flavor observables*, Phys. Rev. D **92**.9 (2015) 094010, DOI: `10.1103/PhysRevD.92.094010`.

[76] E. Maguire, L. Heinrich and G. Watt, *HEPData: a repository for high energy physics data*, J. Phys. Conf. Ser. **898**.10 (2017), ed. by R. Mount and C. Tull 102006, DOI: `10.1088/1742-6596/898/10/102006`.

[77] C. Bierlich *et al.*, *Robust Independent Validation of Experiment and Theory: Rivet version 3*, SciPost Phys. **8** (2020) 026, DOI: `10.21468/SciPostPhys.8.2.026`.

[78] D. Maurin *et al.*, *A cosmic-ray database update: CRDB v4.1*, Eur. Phys. J. C **83**.10 (2023) 971, DOI: `10.1140/epjc/s10052-023-12092-8`.

[79] *LHC Machine*, JINST **3** (2008), ed. by L. Evans and P. Bryant S08001, DOI: `10.1088/1748-0221/3/08/S08001`.

[80] *LHC Design Report Vol.1: The LHC Main Ring* (2004), ed. by O. S. Bruning *et al.*, DOI: `10.5170/CERN-2004-003-V-1`.

[81] CERN, *HL-LHC Industry Project Schedule*, 2025, `https://project-hl-lhc-industry.web.cern.ch/content/project-schedule` (visited on 13/11/2025).

[82] *Longer term LHC schedule*, CERN, 2024, `https://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.html` (visited on 18/11/2025).

[83] K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002, DOI: `10.1088/1748-0221/3/08/S08002`.

[84] B. B. Abelev *et al.*, *Performance of the ALICE Experiment at the CERN LHC*, Int. J. Mod. Phys. A **29** (2014) 1430044, DOI: `10.1142/S0217751X14300440`.

[85] G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003, DOI: `10.1088/1748-0221/3/08/S08003`.

[86] G. Aad *et al.*, *Performance of the ATLAS Detector using First Collision Data*, JHEP **09** (2010) 056, DOI: `10.1007/JHEP09(2010)056`.

[87] S. Chatrchyan *et al.*, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004, DOI: `10.1088/1748-0221/3/08/S08004`.

[88] V. Khachatryan *et al.*, *CMS Tracking Performance Results from Early LHC Operation*, Eur. Phys. J. C **70** (2010) 1165–1192, DOI: `10.1140/epjc/s10052-010-1491-3`.

[89] C. Barschel, *Precision luminosity measurement at LHCb with beam-gas imaging*, Presented 05 Mar 2014, PhD thesis, RWTH Aachen U., 2014, `https://cds.cern.ch/record/1693671`.

[90] E. Mobs, *The CERN accelerator complex - August 2018* (2018), General Photo, `https://cds.cern.ch/record/2636343`.

[91] B. Kniehl *et al.*, *Heavy quark pair production in gluon fusion at next-to-next-to-leading $O(\alpha_s^4)$ order: One-loop squared contributions*, Phys. Rev. D **78** (2008) 094013, DOI: `10.1103/PhysRevD.78.094013`.

[92] R. Aaij *et al.*, *A comprehensive real-time analysis model at the LHCb experiment*, JINST **14**.04 (2019) P04006, DOI: `10.1088/1748-0221/14/04/P04006`.

[93] R. Aaij *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14**.04 (2019) P04013, DOI: `10.1088/1748-0221/14/04/P04013`.

[94] R. Aaij *et al.*, *The LHCb Upgrade I*, JINST **19**.05 (2024) P05065, DOI: `10.1088/1748-0221/19/05/P05065`.

[95] R. Aaij *et al.*, *LHCb SMOG Upgrade*, tech. rep., Geneva: CERN, 2019, DOI: `10.17181/CERN.SAQC.EOWH`, `https://cds.cern.ch/record/2673690`.

[96] R. Antunes-Nobrega *et al.*, *LHCb reoptimized detector design and performance*, Technical design report. LHCb, Geneva: CERN, 2003, `https://cds.cern.ch/record/630827`.

[97] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, DOI: `10.1088/1748-0221/9/09/P09007`.

[98] S. Amato *et al.*, *LHCb magnet*, Technical design report. LHCb, Geneva: CERN, 2000, DOI: `10.17181/CERN.MCVC.Q12G`.

[99] P. R. Barbosa-Marinho *et al.*, *LHCb: Inner tracker technical design report*, 2002, `https://cds.cern.ch/record/582793`.

[100] P. R. Barbosa-Marinho *et al.*, *LHCb: Outer tracker technical design report*, 2001, `https://cds.cern.ch/record/519146`.

[101] R. Arink *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9**.01 (2014) P01002, DOI: `10.1088/1748-0221/9/01/P01002`.

[102] R. Aaij *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. A **30**.07 (2015) 1530022, DOI: `10.1142/S0217751X15300227`.

[103] P. d'Argent *et al.*, *Improved performance of the LHCb Outer Tracker in LHC Run 2*, JINST **12**.11 (2017) P11016, DOI: `10.1088/1748-0221/12/11/P11016`.

[104] M. De Cian *et al.*, *Fast neural-net based fake track rejection in the LHCb reconstruction*, 2017, `https://cds.cern.ch/record/2255039`.

[105] M. Schiller, *Tracking strategies used in LHCb*, 2023, `https://twiki.cern.ch/twiki/bin/viewauth/LHCb/LHCbTrackingStrategies` (visited on 18/11/2025).

[106] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. C **73** (2013) 2431, DOI: `10.1140/epjc/s10052-013-2431-9`.

[107] C. Abellán Beteta *et al.*, *Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC* (2020), `https://cds.cern.ch/record/2729028`.

[108] F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, JINST **8** (2013) P10020, DOI: `10.1088/1748-0221/8/10/P10020`.

[109] A. A. Alves Jr. *et al.*, *Performance of the LHCb muon system*, JINST **8** (2013) P02022, DOI: `10.1088/1748-0221/8/02/P02022`.

[110] M. Ferro-Luzzi, *Proposal for an absolute luminosity determination in colliding beam experiments using vertex detection of beam-gas interactions*, Nucl. Instrum. Meth. **A553** (2005) 388–399, DOI: `10.1016/j.nima.2005.07.010`.

[111] R. Aaij *et al.*, *Precision luminosity measurements at LHCb*, JINST **9**.12 (2014) P12005, DOI: `10.1088/1748-0221/9/12/P12005`.

[112] A. Bursche *et al.*, *Physics opportunities with the fixed-target program of the LHCb experiment using an unpolarized gas target*, tech. rep., Geneva: CERN, 2018, `http://cds.cern.ch/record/2649878`.

[113] C. Lucarelli, *First performance results from upgraded LHCb and SMOG II*, PoS **HardProbes2023** (2024) 196, DOI: `10.22323/1.438.0196`.

[114] S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250, DOI: `10.1016/S0168-9002(03)01368-8`.

[115] J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans.Nucl.Sci. **53** (2006) 270, DOI: `10.1109/TNS.2006.869826`.

[116] S. Acharya *et al.*, *The ALICE definition of primary particles*, 2017, `https://cds.cern.ch/record/2270008`.

[117] R. Aaij *et al.*, *Measurement of Antiproton Production in pHe Collisions at $\sqrt{s_{NN}} = 110$ GeV*, Phys. Rev. Lett. **121**.22 (2018) 222001, DOI: `10.1103/PhysRevLett.121.222001`.

[118] R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10**.02 (2015) P02007, DOI: `10.1088/1748-0221/10/02/P02007`.

[119] M. Schmelling, *Geometric selection of V0 decays*, `https://indico.cern.ch/event/1034187/#2-geometrical-selection-of-v0` (visited on 29/01/2026).

[120] T. Chen and C. Guestrin, *XGBoost: A scalable tree boosting system* (2016).

[121] R. Barlow, *Extended maximum likelihood*, Nucl. Instrum. Meth. **A297** (1990) 496, DOI: `10.1016/0168-9002(90)91334-8`.

[122] Student, *The Probable Error of a Mean*, Biometrika **6**.1 (1908) 1–25, ISSN: 00063444, 14643510, `http://www.jstor.org/stable/2331554` (visited on 15/01/2026).

[123] LHCb Tracking Group, *Long Efficiency Ratios*, `https://twiki.cern.ch/twiki/bin/view/LHCbInternal/LHCbTrackingEfficiencies`, 2019.

[124] T. Skwarnicki, *A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances*, PhD thesis, Cracow, INP, 1986.

[125] H. Dembinski, A. Fedynitch and A. Prosekin, *Chromo: An event generator frontend for particle and astroparticle physics*, Proc. Sci. **ICRC2023** (2023) 189, DOI: `10.22323/1.444.0189`.

[126] J. Adam *et al.*, *HEPData record for the measurement 'Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions'*, HEPData record ins1471838, 2017, `https://doi.org/10.17182/hepdata.77284.v1`.

[127] LHCb Ion Fixed-Target Group, *SMOG2 Monitoring Fill 9523*, 2024, `https://decianm.web.cern.ch/Savesets/fill9523subset/IFT/index.html` (visited on 29/01/2026).

[128] R. Aaij *et al.*, *Measurement of prompt charged-particle production in pp collisions at $\sqrt{s}$ = 13 TeV*, JHEP **01** (2022) 166, DOI: `10.1007/JHEP01(2022)166`.

[129] L. Kolk, *RIVET plug-in for measurement 'Measurement of prompt charged-particle production in pp collisions at* 13 *TeV '*, `https://rivet.hepforge.org/analyses/LHCB_2021_I1889335` (visited on 29/01/2026).

[130] L. Kolk and J. Boelhauve, *Analysis Preservation with HEPData and RIVET*, CERN, LHCb Collaboration, 2023, `https://lhcb-simulation.web.cern.ch/WPG/Rivet/index.html` (visited on 17/12/2025).

[131] R. Aaij *et al.*, *Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions*, Phys. Rev. Lett. **118**.5 (2017), [Erratum: Phys.Rev.Lett. 119, 169901 (2017)] 052002, DOI: `10.1103/PhysRevLett.118.052002`.

[132] R. Aaij *et al.*, *Measurement of B^+, B^0 and Λ_b^0 production in pPb collisions at* $\sqrt{s_{NN}} = 8.16$ *TeV*, Phys. Rev. D **99**.5 (2019) 052011, DOI: `10.1103/PhysRevD.99.052011`.

[133] R. Aaij *et al.*, *Measurement of b hadron fractions in 13 TeV pp collisions*, Phys. Rev. D **100**.3 (2019) 031102, DOI: `10.1103/PhysRevD.100.031102`.

[134] R. Aaij *et al.*, *Study of J/ψ production and cold nuclear matter effects in pPb collisions at* $\sqrt{s_{NN}} = 5$ *TeV*, JHEP **02** (2014) 072, DOI: `10.1007/JHEP02(2014)072`.

[135] R. Aaij *et al.*, *Measurement of the Prompt D^0 Nuclear Modification Factor in p-Pb Collisions at* $\sqrt{s_{NN}} = 8.16$ *TeV*, Phys. Rev. Lett. **131**.10 (2023) 102301, DOI: `10.1103/PhysRevLett.131.102301`.

[136] R. Aaij *et al.*, *Measurement of the Nuclear Modification Factor and Prompt Charged Particle Production in $p-Pb$ and pp Collisions at $\sqrt{s_{NN}}$=5 TeV*, Phys. Rev. Lett. **128**.14 (2022) 142004, DOI: `10.1103/PhysRevLett.128.142004`.

[137] R. Aaij *et al.*, *HEPData record for the measurement 'Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions'*, HEPData record ins1504058, 2024, `https://doi.org/10.17182/hepdata.79130.v1`.

[138] R. Aaij *et al.*, *HEPData record for the measurement 'Measurement of B^+, B^0 and Λ_b^0 production in pPb collisions at* $\sqrt{s_{NN}} = 8.16$ *TeV'*, HEPData record ins1720413, 2024, `https://doi.org/10.17182/hepdata.153895.v1`.

[139] R. Aaij *et al.*, *HEPData record for the measurement 'Measurement of the prompt D^0 nuclear modification factor in pPb collisions at* $\sqrt{s_{NN}} = 8.16$ *TeV'*, HEPData record ins2694685, 2024, `https://doi.org/10.17182/hepdata.153894.v2`.

[140] L. Kolk, J. Boelhauve and G. Daniels, *RIVET plug-in for measurement 'Study of J/ψ production and cold nuclear matter effects in pPb collisions at* $\sqrt{s_{NN}} = 5$ *TeV'*, `https://rivet.hepforge.org/analyses/LHCB_2013_I1251899` (visited on 29/01/2026).

[141] L. Kolk, J. Boelhauve and G. Daniels, *RIVET plug-in for measurement 'Measurement of the b-quark production cross-section in 7 and 13 TeV in pp collisions'*, `https://rivet.hepforge.org/analyses/LHCB_2016_I1504058` (visited on 29/01/2026).

[142] L. Kolk, J. Boelhauve and J. Friedman, *RIVET plug-in for measurement 'Measurement of B^+, B^0 and Λ_b^0 production in pPb collisions at $\sqrt{s_{NN}} = 8.16$ TeV'*, `https://rivet.hepforge.org/analyses/LHCB_2019_I1720413` (visited on 29/01/2026).

[143] L. Kolk, J. Boelhauve and G. Daniels, *RIVET plug-in for measurement 'Measurement of the Nuclear Modification Factor and Prompt Charged Particle Production in $p-Pb$ and pp Collisions at $\sqrt{s_{NN}}$=5 TeV'*, `https://rivet.hepforge.org/analyses/LHCB_2021_I1913240` (visited on 29/01/2026).

[144] L. Kolk, J. Boelhauve and J. Friedman, *RIVET plug-in for measurement 'Measurement of the Prompt D^0 Nuclear Modification Factor in p–Pb Collisions at $\sqrt{s_{NN}} = 8.16$ TeV'*, `https://rivet.hepforge.org/analyses/LHCB_2022_I2694685` (visited on 29/01/2026).

[145] S. Acharya *et al.*, *The ALICE definition of primary particles*, 2017, `https://cds.cern.ch/record/2270008`.

[146] R. Brun and F. Rademakers, *ROOT – An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) 81, DOI: `10.1016/S0168-9002(97)00048-X`.

[147] G. Corti *et al.*, *Software for the LHCb experiment*, IEEE Trans. Nucl. Sci. **53** (2006) 1323, DOI: `10.1109/TNS.2006.872627`.

[148] A. Tsaregorodtsev *et al.*, *DIRAC3: The new generation of the LHCb grid software*, J. Phys. Conf. Ser. **219** (2010) 062029, DOI: `10.1088/1742-6596/219/6/062029`.

[149] E. Rodrigues *et al.*, *The Scikit HEP project - Overview and prospects*, EPJ Web Conf. **245** (2020) 06028, DOI: `10.1051/epjconf/202024506028`.

[150] C. R. Harris *et al.*, *Array programming with NumPy*, Nature **585**.7825 (2020) 357–362, DOI: `10.1038/s41586-020-2649-2`.

[151] The pandas development team, *pandas-dev/pandas: Pandas*, version 2.2.3, 2020, DOI: `10.5281/zenodo.3509134`.

[152] W. McKinney, *Data Structures for Statistical Computing in Python*, *Proceedings of the 9th Python in Science Conference*, ed. by S. van der Walt and J. Millman, 2010 56–61, DOI: `10.25080/Majora-92bf1922-00a`.

[153] J. D. Hunter, *Matplotlib: A 2D graphics environment*, Comput. Sci. Eng. **9** (2007) 90, DOI: `10.1109/MCSE.2007.55`.

[154] H. Dembinski, *HDembinski/jacobi*, 2022, DOI: `10.5281/zenodo.7415783`.

[155] S. K. Lam *et al.*, *numba/numba*, 2023, DOI: `10.5281/zenodo.4343230`.

[156] H. Dembinski, *numba-stats*, 2022, DOI: `10.5281/zenodo.19727013`.

[157] F. Pedregosa *et al.*, *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research **12** (2011) 2825–2830.

[158] P. Virtanen *et al.*, *SciPy 1.0: Fundamental algorithms for scientific computing in Python*, Nat. Meth. **17** (2020) 261, DOI: `10.1038/s41592-019-0686-2`.

[159] F. Mölder *et al.*, *Sustainable data analysis with Snakemake*, F1000Research **10** (2021) 33, DOI: `10.12688/f1000research.29032.2`.

[160] A. Meurer *et al.*, *SymPy: Symbolic computing in Python*, PeerJ Comput. Sci. **3** (2017) e103, DOI: `10.7717/peerj-cs.103`.

[161] J. Pivarski *et al.*, *Awkward Array*, 2018, DOI: `10.5281/zenodo.4341376`.

[162] H. Schreiner *et al.*, *scikit-hep/boost-histogram*, 2022, DOI: `10.5281/zenodo.3492034`.

[163] H. Dembinski *et al.*, *scikit-hep/iminuit*, 2023, DOI: `10.5281/zenodo.3949207`.

[164] E. Rodrigues and H. Schreiner, *scikit-hep/particle*, 2022, DOI: `10.5281/zenodo.2552429`.

[165] J. Pivarski *et al.*, *Uproot*, 2022, DOI: `10.5281/zenodo.4340632`.

[166] S. Chopra *et al.*, *Vector: JIT-compilable mathematical manipulations of ragged Lorentz vectors*, Journal of Open Source Software **10**.109 (2025) 7791, DOI: `10.21105/joss.07791`.

[167] A. Rogozhnikov, *Reweighting with boosted decision trees*, J. Phys. Conf. Ser. **762**.1 (2016) 012036, DOI: `10.1088/1742-6596/762/1/012036`.

Acknowledgement

In this section, I would like to thank everyone who has accompanied and supported me over the last few years while I have been working on this thesis. Firstly, I would like to thank Prof. Dr. Bernhard Spaan, who gave me the opportunity to undertake my doctoral studies in his group. I attended several of his courses and seminars during my Master's degree, and he was very supportive when I approached him about pursuing a doctorate. We talked about the possible topics I could work on, and I liked the idea of working on a topic directly related astroparticle physics. Tragically, he suddenly passed away before I even had the chance to get started.

I would also like to thank Prof. Dr. Johannes Albrecht, who kindly took over as my supervisor after Bernhard Spaan passed away, and welcomed me into his working group. Throughout most of my time at university, he taught several courses and seminars on particle and flavour physics, which strengthened my interest in experimental particle physics. He also supported me when I applied for the CERN summer student programme, one of the best experiences of my life to this day. I would also like to thank Prof. Dr Christian Glaser, who agreed to serve as the second reviewer of my thesis. He has recently joined the university and the SFB 1491 Collaborative Research Centre, where my work was conducted.

Then, I would like to thank Dr. habil. Hans Dembinski, who served as my direct supervisor in our *Astro-QCD* subgroup. He provided me with valuable insights into the Muon Puzzle and the potential of the SMOG system, laying the foundations for my work. I am grateful for all that he taught me about statistics and good programming practices, which I tried my very best to integrate into my analysis code.

The additional support of Prof. Dr. Michael Schmelling is also much appreciated. I enjoyed our weekly meetings, especially the insights you provided on a variety of topics, whether they were related to particle physics, statistics, or something else entirely. It was a pleasure to meet you and collaborate with you, thank you!

After Hans left our group, Dr. Felix Riehn joined and took over as my direct supervisor. Although most of my work had already been completed by that point, I am especially grateful for all our discussions and the insights you provided on air showers and event generators. I am very grateful that you endured my annoying questions and for all the comments you provided as proofreader of my thesis.

An integral part of my time in the Astro-QCD subgroup was spent with Julian, who was a doctoral student when I joined the group. Julian, I very much enjoyed our chats and discussions. You were a great help to me, especially at the beginning when I was transitioning from a rather technical background in top-physics at ATLAS to my first statistical analysis at LHCb. I also enjoyed working with you on the two summer student projects regarding HEPData and RIVET, thank you for everything!

During my time as a doctoral student, I had the pleasure of sharing an office with many people. I want to thank everyone for enduring me, especially Jan, Jonah and Marco, who I shared an office with for the vast majority of time. Many thanks also go to Jamie. I enjoyed our (short) time together in the same office, and am incredibly thankful for the many comments you provided me with during thesis writing, especially considering the time of the year I finalized my thesis in. Obviously, a big shoutout also goes out to Sir Ishaark Newton, the big Blåhaj plush Jamie and I once randomly bought at the close-by Ikea. He has been an office member and unofficial doctoral student ever since. Unfortunately, I have to admit that his made-up thesis title *Search for shark matter at the large hajdron collidor* sounds way cooler than mine. I would also like to thank Luca (also known as Teddy) for all the great times we shared, both inside and outside of university. Thank you also to Noah, Jannis, Luca B., Fran and all the other members of our working group, especially to those taking the time to read my thesis.

Last but not least, I would like to thank my family and friends for their support throughout my journey. Special thanks go to my parents, without which I would not have made it this far. Thank you for providing me a place where I could retreat and switch off every now and then. Thanks also to my little big sister, Liv, who was a great moral support throughout all my ups and downs. I hope this thesis has finally answered your questions about what *exactly* I was doing over the last four years. Thanks also go to Julia for all the time we shared during our bachelor and master studies—especially the lab course that was totally not rushed at all just because someone decided that they needed to go CERN—and everything that came after. And thank you especially for joining me in my new-found hobby of astrophotography, which quite ironically is somehow, in the broadest sense, connected to my actual thesis work.

To all of you: Danke!