

分类号_____

密级_____

UDC _____

编号_____

華中師範大學

博士学位论文

题目：LHC-ALICE 上粲奇异重子 产生的实验研究

学位申请人姓名： 程甜甜

申请学位学生类别： 全日制博士

申请学位学科专业： 粒子物理与原子核物理

指导教师姓名： 殷中宝 教授

Ralf AVERBECK 研究员

Andrea DUBLA 研究员



忠诚博雅，朴实刚毅

Dedicated to my family

谨献给我的家人



博士学位论文
DOCTORAL DISSERTATION



博士学位论文

论文题目：LHC-ALICE 上粲奇异重子 产生的实验研究

论文作者：程甜甜

指导教师：殷中宝 教授

Ralf AVERBECK 研究员

Andrea DUBLA 研究员

学科专业：粒子物理与原子核物理

研究方向：高能核物理实验

华中师范大学物理科学与技术学院

2024 年 5 月



博士学位论文
DOCTORAL DISSERTATION



Dissertation

Measurements of charm-strange baryon production with ALICE at the LHC

By

Tiantian CHENG

Supervisor: Zhongbao YIN, Ralf AVERBECK

Andrea DUBLA

Specialty: Particle Physics and Nuclear Physics

Research Area: High Energy Nuclear Physics

College of Physical Science and Technology

Central China Normal University

GSI Helmholtz Center for Heavy Ion Research

May 2024



华中师范大学学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的研究成果。除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

作者签名：程甜甜

日期：2024年5月30日

学位论文版权使用授权书

学位论文作者完全了解华中师范大学有关保留、使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属华中师范大学。学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许学位论文被查阅和借阅；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。（保密的学位论文在解密后遵守此规定）

保密论文注释：本学位论文属于保密，在____年解密后适用本授权书。

非保密论文注释：本学位论文不属于保密范围，适用本授权书。

作者签名：程甜甜

日期：2024年5月30日

导师签名：

日期：2024年5月30日

本人已经认真阅读“CALIS 高校学位论文全文数据库发布章程”，同意将本人的学位论文提交“CALIS 高校学位论文全文数据库”中全文发布，并可按“章程”中的规定享受相关权益。同意论文提交后滞后：□半年；□一年；□二年发布。

作者签名：程甜甜

日期：2024年5月30日

导师签名：

日期：2024年5月30日



摘要

目前认为，在宇宙大爆炸之后约 10 微秒，宇宙处于极其高温和高致密的物质形态。它经历膨胀、冷却，最终演变成为现在的形态。在它的演变过程中，当系统温度低于化学冻结温度时，核物质会经历一个相变的过程：从解禁闭的状态到夸克胶子束缚在强子内的禁闭状态。根据粒子物理标准模型，量子色动力学（QCD）是用来描写强相互作用的规范理论。QCD 预言，在类似于宇宙早期的极高温和致密的条件下，存在夸克-胶子等离子体（QGP）形态。这是一种强相互作用物质的去约束状态，即夸克和胶子不再受强子的限制。极端高温和致密条件可以在实验室中通过相对论重离子碰撞来达到。大型强子对撞机（LHC）的大型重离子对撞机实验（ALICE）就旨在研究 QGP。

重味，包括粲夸克（c）和底夸克（b），是探索 QGP 的理想硬探针，因为它们产生于碰撞早期的硬散射过程，并能够经历整个碰撞系统的演化过程。高能强子碰撞中粲强子产额的实验测量可以用来检验基于微扰量子色动力学的计算。此外，测量不同类型的强子的产额比值随横动量的分布，可以探索粲夸克的强子化机制。通常认为，QGP 应该不会存在于质子-质子碰撞系统中，但在 TeV 能区的质子-质子碰撞中，已观测到与 Pb-Pb 碰撞中类似的几个现象，如：奇异性增强和重子介子比值的增大。这些实验观测结果挑战着我们目前对于 QGP 产生的认知，因此需要更多的实验数据来进一步探索潜在的物理机制。

包含奇异夸克的重味强子的产生是重离子碰撞实验中一个引人注目的研究课题。因为重夸克可能与热密物质中的其他介质成分组合为强子，与小系统相比，热密介质中奇异夸克对会通过热产生而增多，夸克重新组合机制可能会导致含有奇异夸克的重味强子的产额比那些不含奇异夸克的重味强子要高。该论文旨在通过半轻子衰变道测量粲奇异强子的产额，具体分别是， $\Xi_c^0 \rightarrow \Xi^- e^+ \nu_e$ 和 $\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e$ ，其中， Ξ_c^0 强子含有一个奇异夸克， Ω_c^0 强子含有两个奇异夸克。

Ξ_c^0 强子的产生的测量是基于 LHC-ALICE 实验上采集的 $\sqrt{s} = 5.02$ TeV 的质子-质子碰撞数据。测量的散射截面与已经发表的 7 TeV 和 13 TeV 的结果相比较，表明 Ξ_c^0 强子的产额随着能量的增大而增加。此外，三个能量下粲重子和粲介子产额的比值 Ξ_c^0/D^0 ，都明显地表现出很强的横动量的依赖性。该现象和已测量的 Λ_c^+/D^0 和 $\Sigma^{0,++}/D^0$ 结果相吻合。但是，那些能够很好描述 Λ_c^+/D^0 和 $\Sigma^{0,++}/D^0$ 产额比的理论模型，却无法描述较重的粲奇异强子 Ξ_c^0/D^0 产额比的数据，说明目前还未完全理解粲重子的产生机制。因此，粲奇异重子的实验研究对理论模型有更大的约束力。

目前，通过特定衰变道测量更重粲奇异强子 Ω_c^0 的产生面临的一个大问题是，没有一个精确的绝



对分支比数值。所以，该论文呈现的结果是 Ω_c^0 散射截面乘以其半轻子衰变道的分支比。不过，与之前发表的通过强子衰变道测量的 Ω_c^0 产生的结果相比较，我们可以计算得到这两个衰变道的分支比的比值，即 $BR(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e) / BR(\Omega_c^0 \rightarrow \Omega^- \pi^+) = 1.12 \pm 0.22 \text{ (stat.)} \pm 0.27 \text{ (syst.)}$ 。该测量结果与两个理论计算结果以及 CLEO 合作组的测量结果相吻合，但是却低于 Belle 合作组的测量结果，与之相差 2.3 个标准差。值得注意的是，该 Ω_c^0 分支比的比值和 ALICE 之前发表的 $BR(\Xi_c^0 \rightarrow \Xi^- e^+ \nu_e) / BR(\Xi_c^0 \rightarrow \Xi^- \pi^+)$ 的数值相近。我们期待在 LHC 未来的 Run 3 和 4 中会得到更多更精确的测量结果，这将有助于我们更深入地理解粲夸克的强子化机制和强子碰撞过程中的动力学机制。

关键词：夸克胶子等离子体 (QGP), 大型重离子对撞实验 (ALICE), 质子-质子碰撞, 重味夸克, 粲奇异重子, 强子化, 分支比,



Abstract

Nearly 10^{-6} seconds after the Big Bang, our Universe was in an extremely hot and dense state. It started expanding and cooling down, eventually evolving to its current state. During its evolution, when the temperature drops below the chemical freeze-out temperature, nuclear matter undergoes a phase transition from a deconfined state to a confined hadronic phase, where quarks and gluons are bound together into hadrons. In the Standard Model of particle physics, the gauge theory governing the strong interaction between quarks and gluons is known as quantum chromodynamics (QCD). In conditions of extreme energy density and temperature similar to those in the early Universe, QCD predicts the formation of the quark-gluon plasma (QGP) – a deconfined state of strongly interacting matter, in which quarks and gluons are not confined inside hadrons. These conditions can be reached in the laboratory via relativistic heavy-ion collisions. At the Large Hadron Collider (LHC), A Large Ion Heavy Collider Experiment (ALICE) was the experiment designed to study the QGP.

Heavy flavours, i.e. charm and beauty quarks ($m_c \sim 1.28 \text{ GeV}/c^2$ and $m_b \sim 4.5 \text{ GeV}/c^2$), represent ideal probes of the QGP, since they are produced predominantly via hard-scattering processes in the early stages of the nucleus-nucleus collisions and experience the full system evolution. Measurements of the production of charm hadrons in high-energy hadronic collisions provide important tests for calculations based on perturbative quantum chromodynamics (pQCD). Moreover, investigations of the production ratios of different hadron species as a function of the transverse momentum can shed light on the charm-quark hadronisation mechanism. Generally, the QGP is not expected to exist in small hadronic collision systems, however, enhancement of baryon production, as well as other observed features similar to those in Pb–Pb collisions, such as strangeness enhancement, has also been observed in proton-proton (pp) collisions at the LHC. These phenomena challenge our current understanding of QGP formed only in nucleus-nucleus collisions and of charm hadronisation, therefore, more specific measurements are needed to further explore the underlying physics.

In particular, the production of heavy-flavour hadrons containing strange quarks is an intriguing topic in heavy-ion collisions. As the production of strange quarks is enhanced in the QGP with respect to smaller hadronic collision systems, the possible coalescence of heavy quarks with other medium constituents may lead to an enhanced abundance of heavy-flavour hadrons with strange content compared to those without strangeness. The focus of studies presented in this thesis is on the precise measurements of charmed strange baryons, specifically $\Xi_c^0 \rightarrow \Xi^- e^+ \nu_e$ and $\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e$, where the former contains one strange quark and the latter contains two.



Ξ_c^0 -baryon production is measured with the ALICE detector in pp collisions at a centre-of-mass energy of $\sqrt{s} = 5.02$ TeV. The p_T -differential cross section is compared with previously published measurements at $\sqrt{s} = 7$ and 13 TeV, respectively. The results suggest a hardening of the p_T -differential spectrum with increasing collision energy. The observed p_T dependence of the Ξ_c^0/D^0 yield ratio across the three different collision energies is similar to what was measured for the Λ_c^+/D^0 and $\Sigma_c^{0,++}/D^0$ yield ratios. Model calculations capture the data well for Λ_c^+ and $\Sigma_c^{0,++}$ baryons, however, they fail to describe the heavier strange-charm baryon Ξ_c^0 , indicating that the charm-baryon hadronisation is not fully understood. Therefore, charm-strange baryon measurements have a large constraining power on model predictions.

Currently, the interpretation of the production measurements of even heavier strange-charm Ω_c^0 baryons is limited by the absence of measured absolute branching fraction values. The p_T -differential cross section of Ω_c^0 baryons production multiplied by the branching ratio into $\Omega^- e^+ \nu_e$ is measured in pp collisions at $\sqrt{s} = 13$ TeV in this thesis and divided by existing measurements in the hadronic decay channel $\Omega^- \pi^+$. The resulting branching-fraction ratio $BR(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e) / BR(\Omega_c^0 \rightarrow \Omega^- \pi^+)$ is calculated to be 1.12 ± 0.22 (stat.) ± 0.27 (syst.). The present result is consistent with theory calculations and within one standard deviation with a corresponding measurement from the CLEO Collaboration. However, it is 2.3σ lower than the measurement reported by the BELLE Collaboration. It is also compatible within the uncertainties with the ratio $BR(\Xi_c^0 \rightarrow \Xi^- e^+ \nu_e) / BR(\Xi_c^0 \rightarrow \Xi^- \pi^+)$ measured previously by the ALICE Collaboration. More precise measurements are expected to be carried out in Runs 3 and 4 of the LHC, allowing us to obtain a more complete understanding of charm quark hadronisation mechanisms and dynamics in hadronic collisions.

Keywords: quark–gluon plasma; A Large Ion Collider Experiment (ALICE); proton–proton (pp) collisions; heavy quarks; charm-strange baryon; hadronisation; branching fraction



Contents

摘要	I
Abstract	III
List of Figures	IX
List of Tables	XX
1 General introduction and open heavy flavours	1
1.1 Standard Model	1
1.2 Quantum Chromodynamics	4
1.2.1 Asymptotic freedom	5
1.2.2 Colour confinement	6
1.2.3 Chiral symmetry breaking	6
1.3 Open heavy-flavour physics	7
1.3.1 Open heavy-flavour production in pp collisions	7
1.3.2 Charm quark hadronisation mechanisms	12
1.3.3 Physics motivation	23
1.3.4 Organisation of the thesis	25
2 Experimental setup	27
2.1 The Larger Hadron Collider (LHC)	27
2.2 The ALICE experiment	27
2.2.1 Inner Tracking System (ITS)	29
2.2.2 Time Projection Chamber	31
2.2.3 Time Of Flight	33
2.3 Track and vertex reconstruction	35
2.4 The ALICE offline framework	38
2.5 ALICE detector upgrade	39
2.5.1 ALICE 2 - Run 3	39
2.5.2 ALICE 2 - Run 4	41
2.5.3 ALICE 3 - Run 5 and 6	42
3 Ξ_c^0 production in pp collisions at $\sqrt{s} = 5.02 \text{ TeV}$	43
3.1 Data sample and offline event selection	43



3.2 Raw yield extraction	44
3.2.1 Electron identification	44
3.2.2 Ξ reconstruction	46
3.2.3 $e\Xi$ pairs	47
3.3 Corrections for raw yield extraction	49
3.3.1 Prefilter correction	49
3.3.2 $\Xi_b^{0,-}$ oversubtraction correction	50
3.3.3 Unfolding procedure	53
3.4 Reconstruction efficiency and feed-down subtraction	57
3.5 Systematic uncertainties	60
3.5.1 Track efficiency	60
3.5.2 Electron identification	63
3.5.3 Ξ topology	63
3.5.4 $e\Xi$ pair selection	63
3.5.5 Unfolding procedure	67
3.5.6 Oversubtraction caused by bottom baryons	69
3.5.7 Prompt fraction estimation	69
3.5.8 Generated p_T shape	72
3.5.9 Rapidity range	72
3.5.10 Summary of systematic uncertainties	73
3.6 Results	73
3.6.1 Comparison with model calculations	75
3.6.2 Extrapolation down to $p_T = 0$ of the Ξ_c^0 cross section and the Ξ_c^0/D^0 ratio	76
3.7 Summary	77
4 Ω_c^0 production in pp collisions at $\sqrt{s} = 13$ TeV	79
4.1 Analysis methods	79
4.1.1 Kalman Filter Particle Package	79
4.1.2 Machine Learning	83
4.1.3 Combinatorial background techniques	89
4.2 Candidate selection and reconstruction	90
4.3 Binary classification with XGBoost	94
4.3.1 Training variables	94



4.3.2 Hyperparameters	95
4.3.3 Machine Learning Models	98
4.3.4 Working point determination.....	102
4.4 Signal extraction	104
4.5 Unfolding procedure	107
4.6 Efficiency correction	107
4.7 Systematic uncertainties.....	110
4.7.1 $e\Omega$ -pair selection	110
4.7.2 ITS–TPC matching	114
4.7.3 Track quality selection.....	114
4.7.4 Bayesian-unfolding iterations	116
4.7.5 Unfolding method	117
4.7.6 Response-matrix p_T range and binning.....	117
4.7.7 BDT selection	119
4.7.8 Generated p_T shape.....	120
4.8 Results	121
4.9 Summary	124
5 Summary and outlook	125
5.1 Summary	125
5.2 Outlook.....	127
References	131
Appendix A	145
A.1 Ξ_c^0 analysis	145
A.1.1 Systematic: ITS–TPC matching	145
A.1.2 Systematic: Electron identification	147
A.2 Ω_c^0 analysis	148
A.2.1 Traditional rectangular analysis method	148
A.2.2 Performance with machine learning	153
A.2.3 Performance with mixed event technique	158
A.2.4 Variable distributions and correlation matrix	162
A.2.5 ML output	166



A.2.6 Working Point	170
A.2.7 Acceptance efficiency	173
A.2.8 Systematic: ITS–TPC matching	173
A.2.9 Systematic: BDT selection	177
Appendix B 博士论文中文简介	180
B.1 论文的研究意义	180
B.2 实验分析方法	181
B.2.1 Ξ_c^0 分析方法	182
B.2.2 Ω_c^0 分析方法	182
B.3 实验结果	183
B.4 总结与展望	183
攻读学位期间的成果	185
致 谢	187



List of Figures

1.1	The fundamental particles of the Standard Model of particle physics.	2
1.2	Summary of the measurements of α_s as a function of the energy scale Q . The QCD perturbation theory used to extract the value is listed inside the parenthesis. Figure from [1].	5
1.3	Feynman diagram of leading order quark-antiquark annihilation and gluon fusion for the production of $c\bar{c}$ and $b\bar{b}$. Figure from [2].	8
1.4	Feynman diagram of next-to-leading order, left panel: flavour-excitation, right panel: gluon-splitting. Figure from [2].	9
1.5	Top two panels and bottom left panel: p_T -differential production cross section of D^0 mesons at mid-rapidity ($ y < 0.5$) measured by the ALICE Collaboration in pp collisions at $\sqrt{s} = 5.02$ TeV compared to FONLL [3, 4], GM-VFNS (mod- $\mu_{R,F}$) [5, 6], and k_T -factorisation [7]. Figures from [8]. Bottom right panel: p_T -differential production cross section of B^+ mesons at forward rapidity ($2.0 < y < 4.5$) measured by the LHCb Collaboration in pp collisions at $\sqrt{s} = 7$ TeV compared to FONLL calculations. Figure from [9].	10
1.6	The p_T -differential production cross section for prompt Λ_c^+ in $\sqrt{s} = 5.02$ TeV in pp collisions at ALICE Collaboration [10, 11]. Comparisons to predictions from GM-VFNS [5, 6], POWHEG [12], and PYTHIA 8 [13] are shown, which all significantly underestimate the data at low p_T .	11
1.7	The charm fragmentation fractions $f(c \rightarrow h_c)$ as measured in e^+e^- and e^-p collisions by different experiments [14]. Fractions for pp collisions are shown as well, but these are directly based on the Λ_c^+ production measurement of LHCb [15].	12
1.8	Left panel: ratios between the p_T -differential production cross section of D^+ and D^0 mesons [16]. Right panel: ratios between the D_s^+ -meson and the sum of the D^0 - and D^+ -meson production cross sections. The ratios are compared with prediction obtained with FONLL calculations [3, 4] using PYTHIA 8 [17, 18] for the $H_b \rightarrow D + X$ decay kinematics. For the non-prompt $D_s^+/(D^0 + D^+)$ ratio, the predictions for the D_s^+ from B_s^0 and from non-strange B-meson decays are also displayed separately.	13
1.9	Top left panel: the Λ_c^+/D^0 ratios a function of p_T measured in pp collisions at $\sqrt{s} = 5.02$ TeV [19]. Top right panel: Prompt $\Sigma_c^{0,++}/D^0$ yield ratio in pp collisions at $\sqrt{s} = 13$ TeV [20]. Bottom panel: the $\Xi_c^{0,+}/D^0$ ratios a function of p_T measured in pp collisions at $\sqrt{s} = 13$ TeV [21].	14



1.10 The charm fragmentation fractions at midrapidity ($ y < 0.5$) in pp collisions at $\sqrt{s} = 5.02$ TeV and $\sqrt{s} = 13$ TeV compared with results in e^+e^- and e^-p collisions [14, 22]. The fragmentation fractions $f(c \rightarrow h_c)$ of J/ψ mesons are multiplied by a factor 20 for better visibility.....	16
1.11 Schematic of a multi-parton state with index assignments resulting in a junction-type colour-space ambiguity. Left: The original (LC) string topology. Right: An alternative string topology with a junction and an antijunction, allowed by cyclically matching ‘2’ and ‘5’ indices. Figure from [23].....	18
1.12 The charm quark coalescence probability as a function of the charm quark p_T for pp collisions at LHC. The different coloured lines correspond to the probability of the charm quark to coalesce in a given hadron species, while the black line represents the total coalescence probability. Figure from [24].	21
1.13 Top panel: ratio of the p_T -differential cross section of Ω_c^0 baryons (multiplied by the branching ratio into $\Omega^- \pi^+$) to the D^0 -meson one in pp collisions at $\sqrt{s} = 13$ TeV [25]. Bottom panel: ratio of the data and models to the Catania (coalescence plus fragmentation) model [24].	24
2.1 Schematic layout of the CERN accelerator complex including the LHC with the four big experiments and the injector chain [26].....	28
2.2 The ALICE experimental apparatus with the names of the sub-detectors. Figure from ALICE figure repository.....	29
2.3 Structure of the ITS, Figure from [27].	30
2.4 Layout of the Time Projection Chamber. Figure from [28].	31
2.5 Charged particles dE/dx as a function of momentum measured by the TPC in pp collisions at $\sqrt{s} = 13$ TeV. Figure from ALICE Figure Repository.	32
2.6 Modular structure of the ALICE TOF detector. Figure from [29]	33
2.7 β of charged particles as a function of momentum measured by the TOF in pp collisions at $\sqrt{s} = 13$ TeV. Figure from ALICE Figure Repository.	34
2.8 Event reconstruction flow.	36
2.9 Left panel: reconstruction efficiency for TPC tracks in pp collisions at $\sqrt{s} = 8$ TeV and in central and peripheral Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Right panel: matching efficiency of tracks from ITS to TPC in real (full markers) and simulated (open markers) Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Figures from [29].	36



2.10 Resolution on the primary-vertex position using the SPD and the track algorithms as a function of the charged-particle multiplicity in pp collisions at $\sqrt{s} = 7$ TeV. Figure from [29].	37
2.11 Schematic view of the cross section of the Inner Barrel (left) and Outer Barrel (right). Figure from [30].	40
2.12 Transverse (solid circle markers) and longitudinal (open square markers) impact parameter resolution in Pb–Pb collisions data from ALICE 1 in blue points, and simulations from the upgraded detector ALICE 2 in red points. Figure from [31].	40
2.13 Impact parameter resolution for primary charged pions as a function of the transverse momentum for the ITS3 upgrade and ITS2 upgrade in the transverse plane. The solid lines represent the result with Fast Monte Carlo Tool (FMCT) and ITS only, the dashed lines are the results from FMCT, together with ITS+TPC, and open circles show the results with full MC and ITS only. Figure from [32].	41
3.1 Top left panel: $n\sigma_{\text{TPC}}^{\text{ele}}$ as a function of electron p_T after applying the particle identification criteria on the TOF signal (see text for details). Remaining panels: the distribution of $n\sigma_{\text{TPC}}$ for different p_T^{ele} intervals.	45
3.2 Left panel: the purity of electron with varied $n\sigma_{\text{TPC}}$ selections. Right panel: the contamination of hadrons with varied $n\sigma_{\text{TPC}}$ selections.	46
3.3 Invariant-mass distribution of electron pairs, the red one is with standard cuts and the black one is with prefiltered cuts.	47
3.4 The topological variables are sketched out.	48
3.5 Invariant mass distribution of $\Xi^- \rightarrow \pi^- \Lambda$ (and charge conjugate) candidates integrated over $p_T^{\Xi^-}$. The arrow indicates the world average Ξ^- mass [1] and the dashed lines define the interval in which the Ξ^- candidates are selected for the Ξ_c^0 reconstruction (see text for details). Figure from [33].	49
3.6 Top left panel: the invariant mass distributions of the RS and WS $e\Xi$ pairs. Top right panel: the WS subtracted spectra from RS, raw yield. Bottom left panel: the p_T distributions of the RS and WS $e\Xi$ pairs. Bottom right panel: the ratio between RS and WS spectra as a function of $p_T^{e\Xi}$.	50
3.7 Left panel: Prefilter efficiency ($\epsilon_{e\Xi}$) in different intervals $p_T^{e\Xi}$. Right panel: The comparisons of yield distribution after prefilter correction.	51



3.8 Left panel: the distributions of the beauty-hadron cross sections at $\sqrt{s} = 7$ TeV and 5.02 TeV obtained with FONLL pQCD calculations. Right panel: the fitted Λ_b^0 distribution at $\sqrt{s} = 7$ TeV, and the scaled Λ_b^0 distribution at $\sqrt{s} = 5.02$ TeV.	52
3.9 Left panel: the acceptance-times-efficiency of $\Xi_b^{0,-}$. Right panel: the response matrix of $\Xi_b^{0,-}$ and $e\Xi$	52
3.10 Top left panel: the comparison of distribution of $p_T^{e\Xi}$ from default analysis (black) and from $\Xi_b^{0,-}$ (red). Top right panel: the ratio between red and black, which represents the amount of over-subtraction. Bottom left panel: the comparison of the final result between the WS subtracted spectra without over-subtraction correction (red), and with over-subtraction correction (black). Bottom right panel: the ratio between the corrected and uncorrected result.	53
3.11 Left panel: invariant mass distributions of right-sign and wrong-sign $e\Xi$ pairs with $2 < p_T < 8$ GeV/c. Right panel: invariant mass distribution of Ξ_c^0 candidates obtained by subtracting the wrong-sign pair yield from the right-sign pair yield. Figures from [33]. ..	54
3.12 Left panel: the p_T distribution of Ξ_c^0 in data and in MC, fitted by the <code>expol0</code> function. Right panel: the weighting factor is defined as the ratio of p_T distribution of Ξ_c^0 in data and the in MC.	56
3.13 Left panel: Correlation matrix between the generated Ξ_c^0 -baryon p_T and the reconstructed $e^+\Xi^-$ pair p_T , obtained from the simulation based on PYTHIA 8 described in the text. Figure from [33]. Right panel: Comparisons of the p_T distribution of measured ($p_T^{e\Xi}$), unfolded ($p_T^{\Xi_c^0}$), and refolded ($p_T^{e\Xi}$).	56
3.14 Left panel: product of acceptance and efficiency for prompt and feed-down Ξ_c^0 baryons in pp collisions at $\sqrt{s} = 5.02$ TeV as a function of p_T . Right panel: fraction of prompt Ξ_c^0 baryons in the raw yield (f_{prompt}) as a function of p_T . The systematic uncertainties of f_{prompt} are shown as boxes (explained in Section 3.5.7). Figures from [33].	58
3.15 Top left panel: the Ξ_c^0 spectrum with different selections. Top right panel: ratios of the Ξ_c^0 spectrum with variations. Bottom left panel: Barlow values. Bottom right panel: the root mean square (RMS) of the deviations from 1.	61
3.16 Top left panel: the Ξ_c^0 spectrum with different selections. Top right panel: ratios of the Ξ_c^0 spectrum with variations. Bottom left panel: Barlow value. Bottom right panel: the root mean square (RMS) of the deviations from 1.	62



3.17 The systematic study for electron candidates identification. Top left panel: the Ξ_c^0 spectrum with different selections. Top right panel: ratios of the Ξ_c^0 spectrum with variations. Bottom left panel: Barlow value. Bottom right panel: the root mean square (RMS) of the deviations from 1.	64
3.18 The systematic study for Ξ reconstruction. Top left panel: the Ξ_c^0 spectrum with different selections. Top right panel: ratios of the Ξ_c^0 spectrum with variations. Bottom left panel: barlow value. Bottom right panel: the root mean square (RMS) of the deviations from 1.	65
3.19 The systematic study for $e\Xi$ pair selection. Left panel: the Ξ_c^0 spectra with different selections. Right panel: ratios of the Ξ_c^0 spectra with variations.	66
3.20 The systematic study for the unfolding procedure of varied Bayesian iteration numbers. Top left panel: the p_T distribution of unfolded yield Ξ_c^0 with different Bayesian iteration numbers. Top right panel: ratios of the unfolded yield Ξ_c^0 with variation. Bottom panel: the root mean square (RMS) of the deviations from 1.	67
3.21 The systematic study for the unfolding procedure of Singular Value Decomposition (SVD). Top left panel: the comparison of p_T distribution of unfolded yield Ξ_c^0 from the SVD method compared to the default Bayesian one. Top right panel: ratios of the unfolded yield Ξ_c^0 with variation. Bottom panel: the root mean square (RMS) of the deviations from 1.	68
3.22 The systematic study for the unfolding procedure of p_T binning. Left panel: the p_T distribution of unfolded yield Ξ_c^0 with different p_T binning. Right panel: ratios of the unfolded yield Ξ_c^0 with variation.	69
3.23 The systematic study for oversubtraction by varying the parameters of the TSallis functions.	70
3.24 The systematic study for oversubtraction by varying branching ratio fraction.	70
3.25 The systematic study for prompt fraction estimation.	71
3.26 The systematic study for prompt fraction estimation. Left panel: The comparisons of prompt fractions obtained from different assumptions. Right panel: The ratio of prompt fractions between the varied ones over the standard one.	71
3.27 The systematic study for the generated p_T shape.	72
3.28 The systematic study for rapidity range selection.	72



3.29 Left panel: p_T -differential production cross sections of prompt Ξ_c^0 baryons in pp collisions at $\sqrt{s} = 5.02$ TeV and 13 TeV [21] and of inclusive Ξ_c^0 baryons in pp collisions at $\sqrt{s} = 7$ TeV [34] with updated decay BR as discussed in the text. The uncertainty of the BR of the cross sections of prompt Ξ_c^0 baryons in pp collisions at $\sqrt{s} = 13$ TeV is lower because it consists in the combination of two different decay channels ($\Xi_c^0 \rightarrow e^+ \Xi^- \nu_e$ and $\Xi_c^0 \rightarrow \pi^+ \Xi^-$) [21]. Right panel: Ξ_c^0/D^0 ratio measured in pp collisions at $\sqrt{s} = 5.02$ TeV, compared with the measurements at $\sqrt{s} = 7$ TeV [34] and $\sqrt{s} = 13$ TeV [21]. The uncertainty of the BR of D^0 and Ξ_c^0 are shown as shaded boxes. Figures from [33].	74
3.30 Left panel: p_T -differential production cross section of prompt Ξ_c^0 baryons in pp collisions at $\sqrt{s} = 5.02$ TeV compared with model calculations [23, 35, 36]. Right panel: Ξ_c^0/D^0 ratio as a function of p_T measured in pp collisions at $\sqrt{s} = 5.02$ TeV compared with model calculations [23, 24, 35–37] (see text for details).	76
4.1 Sketch of the semileptonic decay of a Ω_c^0 baryon. The primary vertex (PV) is the collision point where the Ω_c^0 baryon is produced.	83
4.2 Example of the separation between signal (blue) and background (red). the black line fits the data well, the green line is overfit.	85
4.3 Example of a decision tree structure for binary classification of signal (S , in green) and background (B , in red) candidates. The initial sample at the first node is split into branches. The terminal nodes are shown as circles and are either signal leaves in the case that the signal candidates are dominant, and vice versa for the background leaves.	86
4.4 Schematic of a boosting procedure.	87
4.5 Comparison of invariant mass distribution of $e\Omega$ pairs between SE and ME in three p_T intervals.	90
4.6 Comparison of invariant mass distribution of $e\Omega$ pairs between SE and TR in three p_T intervals.	90
4.7 Comparison of invariant mass distribution of $e\Omega$ pairs in the interval $2 < p_T^{e\Omega} < 12$ GeV/ c .	91
4.8 Training variables (expect EleOmegaMass, EleOmegapT, CosOA) distribution in $2 < p_T < 4$ for signal (reported in orange) and background (reported in blue) candidates.	96
4.9 Correlation matrix in the transverse momentum range $2 < p_T < 4$.	96
4.10 The learning curves of the training set and test set in $2 < p_T < 4$ GeV/ c .	99
4.11 The ROC-AUC curves for the training set and test set in $2 < p_T < 4$ GeV/ c .	100
4.12 Machine learning model in $2 < p_T < 4$ GeV/ c .	100
4.13 Feature importance of the classification criteria in $2 < p_T < 4$ GeV/ c .	101



4.14 Left panel: The number of background candidates (MERS scaled by SEWS) as a function of BDT cut for $2 < p_T < 4 \text{ GeV}/c$. Right panel: The number of estimated expected signal candidates as a function of BDT cut $2 < p_T < 4 \text{ GeV}/c$.	103
4.15 The pseudo-significance as a function of BDT cut for $2 < p_T < 4 \text{ GeV}/c$.	103
4.16 The invariant mass distribution of $e\Omega$ for RS and WS in SE, and MERS, at each p_T interval.	104
4.17 The mass distribution comparison: SERS-MERS and SERS-SEWS, at each p_T interval.	105
4.18 Left panel: invariant-mass distribution of opposite-sign pairs (black solid circle marker) and same-sign pairs (red solid square marker) in SE, and opposite-sign pairs (blue open square marker) in ME. Right panel: invariant-mass distribution of the Ω_c^0 candidates obtained by subtracting the opposite-sign charge $e\Omega$ pairs in ME from the opposite-sign charge pairs in SE (black solid circle marker), and $e\Omega$ opposite-sign charge pairs coming from Ω_c^0 decay from PYTHIA 8 (green open circle marker).	105
4.19 Left: The prefilter efficiency. Right: The comparison of yield distribution of $p_T e\Omega$.	106
4.20 Left panel: The correlation matrix between the generated Ω_c^0 -baryon p_T and the reconstructed $e\Omega$ pair p_T after weights applied. Right panel: Comparisons of the measured spectrum to the refolded and the unfolded one with and without weights.	107
4.21 Top left panel: (Weighted) Preselection efficiency. Top right panel: (Weighted) BDT efficiency. Bottom panel: (Weighted) Total efficiency.	109
4.22 Invariant-mass distribution of the $e\Omega$ pairs from PYTHIA 8 simulation study in different p_T intervals.	112
4.23 Left panel: Comparison of the Ω_c^0 spectrum with and without the mass cut ($M_{e\Omega} > 2\text{GeV}/c^2$) applied. Right panel: the ratio between the Ω_c^0 spectrum with mass cut over Ω_c^0 spectrum without one.	113
4.24 Top left panel: the Ω_c^0 spectrum with different mass selections. Top right panel: ratios of the Ω_c^0 spectrum with variations. Bottom left panel: Barlow value. Bottom right panel: the root mean square (RMS) of the deviations from 1.	113
4.25 Top left panel: The Ω_c^0 spectrum with different track selection criteria. Top right panel: The ratios. Bottom left panel: The Barlow value for each variation. Bottom right panel: The RMS of the deviation.	115
4.26 Top left panel: the comparisons of the distribution of unfolded yields with the different variations. Top right panel: the ratios of the spectra with different unfolding methods relative to the spectrum with the standard procedure. Bottom left panel: the Barlow values for each variation. Bottom right panel: the rms values of the deviation from 1.	116



4.27	The distribution of d vector.	117
4.28	Top left panel: the comparisons of the distribution of unfolded yields with the different variations. Top right panel: the ratios of the spectra with different unfolding methods relative to the spectrum with the standard procedure. Bottom left panel: the Barlow values for each variation. Bottom right panel: the rms values of the deviation from 1.	118
4.29	Left panel: the comparisons of the distribution of unfolded yields with the different p_T intervals. Right panel: ratios of the varied distributions with different p_T intervals to the default one.	118
4.30	Left panel: the comparison of the Ω_c^0 spectrum with variations of the p_T dependent $n\sigma_{\text{TPC}}^e$. Right panel: the ratios of the Ω_c^0 spectrum with variations.	119
4.31	Top left panel: the Ω_c^0 spectrum distribution between the Generated and hadronic result [25], and also three different Tsallis fitting to the hadronic result. Top right panel: the three weighting functions. Bottom left panel: comparisons of the reconstruction efficiency after re-weighting procedure with three different weighting functions. Bottom right panel: ratios of the reconstruction efficiency efficiency between the varied Tsallis fitting functions to the central one.	120
4.32	Top panel: p_T -differential production cross sections of inclusive Ω_c^0 baryons multiplied by the branching ratios into $\Omega^- e^+ \nu_e$ and $\Omega^- \pi^+$ [25] in pp collisions at $\sqrt{s} = 13$ TeV. Bottom panel: p_T -differential branching-fraction ratio $\text{BR}(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e)/\text{BR}(\Omega_c^0 \rightarrow \Omega^- \pi^+)$	121
4.33	Comparison of $\text{BR}(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e)/\text{BR}(\Omega_c^0 \rightarrow \Omega^- \pi^+)$ between experimental and theoretical calculations [38–42].	123
5.1	Top panels: Invariant-mass distribution of baryons in pp collisions at $\sqrt{s} = 13.6$ TeV. The background distribution is modeled with a second-order polynomial, while the signal with a Gaussian function. Left panel: The invariant mass distribution of baryon Λ_c^+ with $6 < p_T < 8$ GeV/c. Right panel: The invariant mass distribution of Ξ_c^0 candidates from hadronic decay channel $\Xi^- \pi^+$ with $p_T > 5$ GeV/c. Bottom panel: Yield ratio of Λ_b/B^+ as a function of p_T in Pb–Pb collisions at $\sqrt{s_{\text{NN}}} = 5.52$ TeV. Simulation for Run 3 + Run 4. Figures from [43].	128
5.2	Expected Ξ_{cc}^{++} and Ω_{cc}^+ significance in 0-10% central Pb–Pb collisions at $\sqrt{s_{\text{NN}}} = 5.52$ TeV as a function of p_T with a 0.2 T magnetic field. Figures from [43].	129
A.1	The $p_T^{e \leftarrow \Xi_c^0}$ distributions in different $p_T^{e \Xi}$ intervals.	146



A.2	The $p_T^{e\Xi}$ distributions in different $p_T^{\Xi_c^0}$ intervals.	147
A.3	The varied criteria of the TPC dE/dx signal to select candidates, $n\sigma_{\text{TPC}}^{\text{ele}}$, is applied in this systematic study.	148
A.4	The invariant mass distribution of $e\Omega$ pairs for RS and WS in SE in each p_T interval.	151
A.5	The invariant mass distribution of $e\Omega$ pairs after the WS mass subtraction in each p_T interval.	151
A.6	Left: The prefilter efficiency. Right: The comparison of yield distribution of $p_T e\Omega$.	152
A.7	Left: The correlation matrix between the generated Ω_c^0 -baryon p_T and the reconstructed $e\Omega$ pair p_T . Right: Comparisons of the measured spectrum to the refolded and the unfolded one with and without weights.	152
A.8	Left panel: product of acceptance and efficiency for prompt, feed-down, and inclusive Ω_c^0 baryons. Right panel: p_T -differential production cross sections of inclusive Ω_c^0 baryons multiplied by the branching ratio into $\Omega^- e^+ \nu_e$.	153
A.9	Machine learning model in each p_T interval in this analysis when SEWS is used for the background. Top left panel: $2 < p_T < 4 \text{ GeV}/c$. Top right panel: $4 < p_T < 6 \text{ GeV}/c$. Bottom panel: $6 < p_T < 12 \text{ GeV}/c$.	153
A.10	The pseudo-significance in each p_T interval.	154
A.11	The invariant-mass distribution of $e\Omega$ RS and WS pairs in each p_T interval.	155
A.12	The invariant-mass distribution of $e\Omega$ pairs after WS subtracting from RS pairs in each p_T interval.	155
A.13	Left panel: The response matrix between the generated Ω_c^0 -baryon p_T and the reconstructed $e\Omega$ pair p_T . Right panel: The raw yield comparison.	156
A.14	Top left: Preselection efficiency. Top right: BDT efficiency. Bottom: total acceptance times efficiency for the inclusive Ω_c^0 baryons.	156
A.15	Left panel: The comparison of p_T -differential production cross section of inclusive Ω_c^0 baryons multiplied by the branching ratio into $\Omega^- e^+ \nu_e$, between the traditional rectangular cut method (in red color) and the new ML method (in black color). Right panel: The improvement of relative statistical uncertainties from ML with respect to the traditional rectangular cut method.	157
A.16	The comparison of variables between MERS (reported in blue) and SEWS (reported in orange), in $2 < p_T < 4$.	158
A.17	The comparison of variables between MERS (reported in blue) and SEWS (reported in orange), in $4 < p_T < 6$.	159



A.18 The comparison of variables between MERS (reported in blue) and SEWS (reported in orange), in $2 < p_T < 4$.	159
A.19 Left panel: The raw yield comparison from two different ML models: MERS and SEWS. Right panel: The raw yield ratio.	160
A.20 Left panel: The total reconstruction efficiency comparison from two different ML models: MERS and SEWS. Right panel: The total reconstruction efficiency ratio.	160
A.21 Left panel: The comparison of p_T -differential production cross section of inclusive Ω_c^0 baryons multiplied by the branching ratio into $\Omega^- e^+ \nu_e$ from two different ML models: MERS and SEWS. Right panel: The p_T -differential production cross section ratio.	160
A.22 Training variables (expect EleOmegaMass, EleOmegapT, CosOA) distribution in $1 < p_T < 2$ for signal (reported in orange) and background (reported in blue) candidates.	162
A.23 Training variables (expect EleOmegaMass, EleOmegapT, CosOA) distribution in $4 < p_T < 6$ for signal (reported in orange) and background (reported in blue) candidates.	163
A.24 Training variables (expect EleOmegaMass, EleOmegapT, CosOA) distribution in $6 < p_T < 12$ for signal (reported in orange) and background (reported in blue) candidates.	163
A.25 Correlation matrix in the transverse momentum range $1 < p_T < 2$.	164
A.26 Correlation matrix in the transverse momentum range $4 < p_T < 6$.	164
A.27 Correlation matrix in the transverse momentum range $6 < p_T < 12$.	165
A.28 The learning curves of the training set and test set in each p_T interval in this analysis. Top left panel: $1 < p_T < 2$ GeV/c. Top right panel: $4 < p_T < 6$ GeV/c. Bottom panel: $6 < p_T < 12$ GeV/c.	166
A.29 The ROC-AUC curves for the training set and test set in each p_T interval in this analysis. Top left panel: $1 < p_T < 2$ GeV/c. Top right panel: $4 < p_T < 6$ GeV/c. Bottom panel: $6 < p_T < 12$ GeV/c.	167
A.30 Machine learning model in each p_T interval in this analysis. Top left panel: $1 < p_T < 2$ GeV/c. Top right panel: $4 < p_T < 6$ GeV/c. Bottom panel: $6 < p_T < 12$ GeV/c.	168
A.31 Feature importance of the classification criteria in each p_T interval in this analysis. From top to bottom: $1 < p_T < 2$ GeV/c, $4 < p_T < 6$ GeV/c, and $6 < p_T < 12$ GeV/c.	169
A.32 The number of background candidates (MERS scaled by SEWS) as a function of BDT cut for each p_T interval in this analysis. Top left panel: $1 < p_T < 2$ GeV/c. Top right panel: $4 < p_T < 6$ GeV/c. Bottom panel: $6 < p_T < 12$ GeV/c.	170



A.33 The number of estimated expected signal candidates as a function of BDT cut for each p_T interval. Top left panel: $1 < p_T < 2 \text{ GeV}/c$. Top right panel: $4 < p_T < 6 \text{ GeV}/c$. Bottom panel: $6 < p_T < 12 \text{ GeV}/c$.	171
A.34 The pseudo-significance as a function of BDT cut for each p_T interval. Top left panel: $2 < p_T < 4 \text{ GeV}/c$. Top right panel: $4 < p_T < 6 \text{ GeV}/c$. Bottom panel: $6 < p_T < 12 \text{ GeV}/c$.	172
A.35 $1 < p_T < 2$: The Tsallis function is used to fit the result of hadronic Ω_c^0 . The pseudo-significance as a function of BDT cut ($WP_{BDT} = 0.95$) to reject backgrounds from other hadrons.	172
A.36 Product of acceptance and efficiency for prompt, feed-down, and inclusive Ω_c^0 baryons in this analysis.	173
A.37 The distributions of p_T^e in different $p_T^{e\Omega}$ intervals.	174
A.38 The distributions of $p_T^{e\Omega}$ in different $p_T^{\Omega_c^0}$ intervals.	175
A.39 Top left panel: the ratios of the raw yield as a function of the BDT cuts to the default one. Top right panel: the ratios of the reconstruction efficiency as a function of the BDT cuts to the default one. Bottom left panel: the ratios of the corrected yield as a function of the BDT cuts to the default one. Bottom right panel: the systematic uncertainty, $\sqrt{(RMS)^2 + (shift)^2}$, as a function of BDT in $2 < p_T < 4$.	177
A.40 Top left panel: the ratios of the raw yield as a function of the BDT cuts to the default one. Top right panel: the ratios of the reconstruction efficiency as a function of the BDT cuts to the default one. Bottom left panel: the ratios of the corrected yield as a function of the BDT cuts to the default one. Bottom right panel: the systematic uncertainty, $\sqrt{(RMS)^2 + (shift)^2}$, as a function of BDT in $4 < p_T < 6$.	178
A.41 Top left panel: the ratios of the raw yield as a function of the BDT cuts to the default one. Top right panel: the ratios of the reconstruction efficiency as a function of the BDT cuts to the default one. Bottom left panel: the ratios of the corrected yield as a function of the BDT cuts to the default one. Bottom right panel: the systematic uncertainty, $\sqrt{(RMS)^2 + (shift)^2}$, as a function of BDT in $6 < p_T < 12$.	179



List of Tables

1.1	Thermal densities of “prompt” ground-state charmed hadrons for hadronization temperatures of $T_H = 170$ MeV (including strong feeddowns) in the PDG and also added the one expected by the RQM. Numbers from [37].	22
2.1	Brief summary of the different layers of the ALICE ITS.	29
3.1	Electron identification cuts applied in this analysis.	44
3.2	Properties of Ξ . The valence quark content, the strangeness S, the isospin, the spin and parity $I(J^P)$, the mass and the main decay channel, together with its branching ratio (BR) and decay length $c\tau$ are listed.	47
3.3	The track and topology selections for Ξ applied in this analysis.	48
3.4	Varied criteria for electron selections. Selections that are not shown in this table are the same as the ones in Tab. 3.1.	60
3.5	Ξ track cuts used to study the uncertainties in Ξ reconstruction. Other cuts are not shown which are the same as in Tab. 3.3.	62
3.6	The systematic uncertainty of $p_T^{\Xi_c^0}$ propagated from $p_T^{e\Xi}$	63
3.7	Eletron selections used to study the uncertainties of track identification. Cuts that are not shown in this table are the same as the ones in Tab. 3.1.	66
3.8	The varied selections of Ξ topology to study the uncertainties for the daughter track reconstruction. Selections that are not shown in this table are the same as Tab. 3.3.	66
3.9	The variation of pairs mass selection.	66
3.10	Contributions to the systematic uncertainty of the prompt fraction estimation.	71
3.11	Contributions to the systematic uncertainty of the Ξ_c^0 cross section in this analysys.	73
4.1	Tracking cuts for electrons applied in this analysis.	92
4.2	The tracking cuts for Ω daughters applied in this analysis.	92
4.3	The further pre-selections applied in this analysis.	92
4.4	Number of signal and background candidates used for the BDT model training and testing in the analysed p_T intervals.	94
4.5	Optimised hyperparameters used in the analysed for each p_T interval.	97
4.6	Determined working point values used for each p_T intervals in this analysis.	103



4.7 Contributions to the systematic uncertainty of the Ω_c^0 cross section for the p_T intervals $2 < p_T < 4 \text{ GeV}/c$, $4 < p_T < 6 \text{ GeV}/c$, and $6 < p_T < 12 \text{ GeV}/c$	110
4.8 The cut variation on $M_{e\Omega}$ (GeV/c^2).....	113
4.9 The systematic uncertainty of $p_T^{\Omega_c^0}$ propagated from p_T^e	114
4.10 The variations of TPC selection criteria in the pre-selection used for this systematic study.....	114
A.1 ITS-TPC matching efficiency for electron taken from DPG [44]	145
A.2 The systematic uncertainty of ITS-TPC matching efficiency of $p_T^{e\Xi}$ propagated from p_T^e	146
A.3 The systematic uncertainty of $p_T^{\Xi_c^0}$ propagated from $p_T^{e\Xi}$	148
A.4 eID cuts applied in this analysis.....	149
A.5 Loose eID cuts applied for the prefilter procedure in this analysis.....	149
A.6 The track and topology selections for Ω applied in this analysis.....	150
A.7 The relative statistics error of the p_T -differential production cross section of inclusive Ω_c^0 baryons multiplied by the branching ratio into $\Omega^- e^+ \nu_e$ from two different ML models shown in Fig. A.21, for the p_T intervals $2 < p_T < 4 \text{ GeV}/c$, $4 < p_T < 6 \text{ GeV}/c$, and $6 < p_T < 12 \text{ GeV}/c$	161
A.8 The systematic uncertainties of ITS-TPC matching efficiency for electron taken from DPG in each p_T bin at $\sqrt{s} = 13 \text{ TeV}$	173
A.9 The systematic uncertainty of $p_T^{e\Omega}$ propagated from p_T^e in this analysis.....	174
A.10 The systematic uncertainty of $p_T^{\Omega_c^0}$ propagated from $p_T^{e\Omega}$	176
A.11 Electron identification used to study the uncertainties in eID.....	177



博士学位论文
DOCTORAL DISSERTATION





1 General introduction and open heavy flavours

The strong interaction among quarks and gluons, the elementary constituents of hadronic matter, is described by quantum chromodynamics (QCD). QCD is a fundamental component of the Standard Model (SM) in particle physics. One of the fundamental properties of QCD is the dependence of the strength of the interactions between two colour charges on the squared 4-momentum transfer (Q^2) involved in the process. The interactions between colour-charged quarks and gluons become weaker with increasing Q^2 . This leads to the property of “color confinement,” where the potential between color-charged particles rises linearly with their separation. Consequently, quarks and gluons cannot exist in isolation but are confined within composite objects known as hadrons, which are observed in the final state of high-energy particle collisions. In this chapter, a brief introduction to the SM is presented and then followed by a description of the QCD theory. Open heavy-flavour hadrons are among the most important tools for studying QCD in high-energy hadronic collisions, and in particular their production in proton–proton (pp) collisions can be used as a stringent test of perturbative QCD (pQCD) calculations. Additionally, several selected theoretical models about hadronisation are shown, with particular emphasis on heavy-flavour baryon production aspects.

1.1 Standard Model

The matter surrounding us interacts, in our current understanding, through four fundamental forces. These forces, listed in ascending order of their coupling strength, include gravitational, weak, electromagnetic, and strong interactions. The interactions among elementary particles of matter (fermions) occur through the exchange of force carriers (gauge bosons) and are described by the SM. Fig. 1.1 illustrates all the bosons and fermions introduced in the SM. According to their spin, they are grouped into fermions, which are half-integer spin matter particles, and bosons, which are integer spin force carriers mediating SM interactions. Fermions are further divided into two groups: quarks and leptons. Quarks can be of six different flavours: *up*, *down*, *strange*, *charm*, *beauty* and *top*. In the following, the four fundamental interactions and their relation to the elementary particles of the SM are introduced.

Gravity is a force of attraction between two objects carrying mass and energy. It is the first force to be fully understood at a large distance, but it is likely to be the last to be understood at very short distances. This is because the coupling strength of gravity is very weak compared to the other three interactions, with a coupling constant that is 10^{34} times smaller than the electromagnetic coupling constant α_{EM} . Gravitational effects are responsible for many astronomical phenomena, such as the formation of stars



Standard Model of Elementary Particles

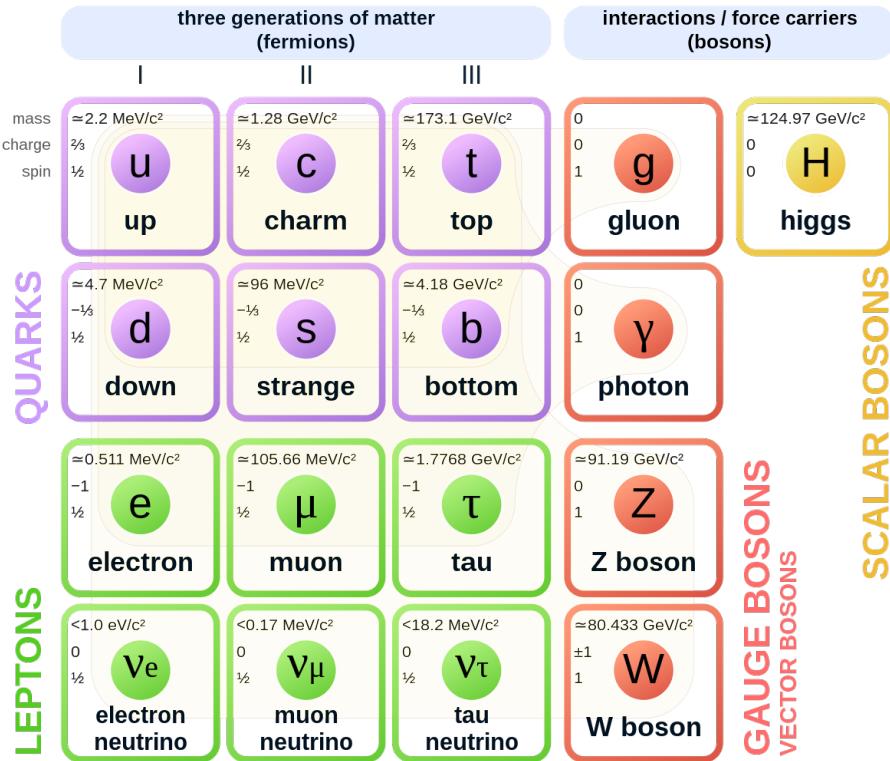


Figure 1.1: The fundamental particles of the Standard Model of particle physics.

and the trajectories of astronomical bodies. However, gravitational effects would not be observed in particle collisions below centre-of-mass energy close to the Planck scale (10^{19} GeV), which is well beyond the energy limit of current particle colliders. The gravitational interaction is described by the general theory of relativity, which considers a geometric property of a four-dimensional space-time. It is not included in the SM, as general relativity is a classical theory. A consistent and predictive quantum theory of gravity, often called a “quantum theory of gravity” has not been formulated yet.

The electromagnetic interaction, which describes the forces between electrically charged particles, is mediated by the exchange of photons. This interaction includes both the electrostatic force and a combination of electric and magnetic forces (magnetism). The electrostatic force occurs between any two charged particles, leading to attraction between particles with opposite charges and repulsion between particles with the same charge. On the other hand, the magnetic force arises exclusively between charged particles in relative motion. Notably, the electromagnetic force has an infinite range, due to the massless nature of the photon. Its strength decreases with increasing distance between particles ($\sim 1/r$ for the electro-



magnetic potential), as explained by the coupling constant α_{EM} in Eq. 1.1 [45], commonly known as the fine structure constant:

$$\alpha_{\text{EM}} = \frac{1}{4\pi\epsilon_0 \hbar c} \frac{e^2}{137.035999084}, \quad (1.1)$$

where $e \approx 1.602 \cdot 10^{-19}$ C is the magnitude of the electron charge, $\epsilon_0 \approx 8.854 \cdot 10^{-12}$ F/m represents the permittivity of free space, $\hbar \approx 1.054 \cdot 10^{-34}$ J · s is the reduced Planck constant, and $c \approx 2.998 \cdot 10^8$ m/s is the speed of light in vacuum [45]. This energy-scale dependence of the coupling constant is known as a running coupling. The electromagnetic interaction is successfully described by the theory of Quantum Electrodynamics (QED).

The weak interaction describes the underlying mechanism behind the decay of some unstable particles, such as the β decay of radioactive nuclei. This force acts on all fermions, as well as the Higgs boson. It is mediated by its three force carriers, the massive W^+ , W^- , and Z^0 bosons. Because the force carriers are massive, the weak interaction has an extremely short effective range, in the order of $\sim 10^{18}$ m. The coupling strength of the weak interaction is approximately 10^3 times smaller than that of α_{EM} in Eq. 1.1. One distinctive feature of the weak interaction is its ability to allow quarks to change their flavor. At high energies or short distances, the strength of the weak interaction becomes comparable to that of the electromagnetic interaction. Both interactions can be unifiedly described by the electroweak theory.

Closely related to the electroweak theory is the Higgs field, which plays a key role in the Higgs mechanism [46, 47]. It is responsible for marking three of the boson fields with mass, and becoming the known weak bosons of physics: W^+ , W^- , Z^0 , as well as assigning masses to the charged leptons and quarks through the Yukawa interaction [48]. The Higgs mechanism predicts the existence of a complex scalar boson, the Higgs boson, which was the last addition to complete the picture of the SM. The Higgs boson was successfully discovered by the ATLAS and CMS experiments at the LHC in CERN, Geneva in 2012 [49, 50].

The strong interaction is responsible for binding quarks and gluons together, forming composite particles such as mesons and baryons, as well as other exotic states like the recently observed tetraquarks and pentaquarks. Following the observation of the three up quark systems, Δ^{++} consisting of three quarks of the same flavour in 1952 [51], the colour quantum number was introduced as an extra degree of freedom in the quark model. This addition prevents quarks in baryons from violating the Pauli exclusion principle. The strong force, mediated by the gluons, acts on the colour charge carried by quarks and gluons. Each quark is assigned with one of three colour: red (r), blue (b) or green (g), while anti-quarks have their anti-colour (anti-red \bar{r} , anti-blue \bar{b} , or anti-green \bar{g}). Gluons themselves carry a combination of colour



and anti-colour. At the typical length scale of hadrons (a few femtometers), the coupling constant of the strong interaction is on the order of 1. The strong interaction is described within the framework quantum chromodynamics framework (QCD), discussed in Section 1.2.

1.2 Quantum Chromodynamics

The strong interaction between quarks and gluons is formulated as a non-abelian gauge field theory called QCD based on the invariance under $SU(3)_c$ group transformations. It was developed in analogy to quantum electrodynamics (QED). The QCD's name refers to the concept of *colour charge*, which is a quantum property of quarks. In QCD, quarks interact with each other through the exchange of gluons, similarly to electrically charged particles interacting via the exchanges of photons in QED.

The QCD Lagrangian describing the dynamics of quarks and gluons is given by

$$\mathcal{L}_{QCD} = \sum_{f=1}^{n_f} \bar{q}_i^f \left(i\gamma_\mu D_{ij}^\mu - m_f \delta_{ij} \right) q_j^f - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu}, \quad (1.2)$$

where q_i^f is the quark field spinor of a quark with flavour f (u, d, c, s, t, b), and a vector index i : $q_{\text{red}}, q_{\text{blue}}, q_{\text{green}}$ that runs from 1 to 3. The term $-m_f \bar{q} q$ explicitly includes the mass of the quarks, which are generated by the Yukawa interaction[48]. The gauge covariant derivative D_{ij}^μ is defined as,

$$D_{ij}^\mu = \partial^\mu \delta_{ij} - ig_s \left(\frac{\lambda_{ij}^a}{2} \right) G_a^\mu, \quad (1.3)$$

where γ^μ is the Dirac matrices, and δ_{ij} is the Dirac delta function. The G_a^μ corresponds to the gluon fields with $a = 1, \dots, 8$. The gluon fields do not distinguish between the quark flavours. The λ_{ij}^a are 3×3 Gell-Mann matrices.

The gluon field strength tensor $G_{\mu\nu}^a$, the second term in Eq. 1.2, encodes the interaction of gluons with the quarks and the gluon self-interaction. It is given by

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + ig_s f_{abc} G_\mu^b G_\nu^c, \quad (1.4)$$

where the a, b , and c are the colour charges in the $SU(3)$ adjoint representation, f_{abc} are the structure constants of $SU(3)$, and g_s indicates the intensity of the strong interaction. The non-Abelian term, $ig_s f_{abc} G_\mu^b G_\nu^c$, leads to the self-interactions among gluons, resulting in what is called *anti-screening* in colour interaction. Gluons can decay into new gluon pairs, a consequence of the possibility of gluon self-interaction.

QCD exhibits three distinctive features: 1) Asymptotic freedom, 2) Colour confinement and 3) Chiral symmetry breaking.

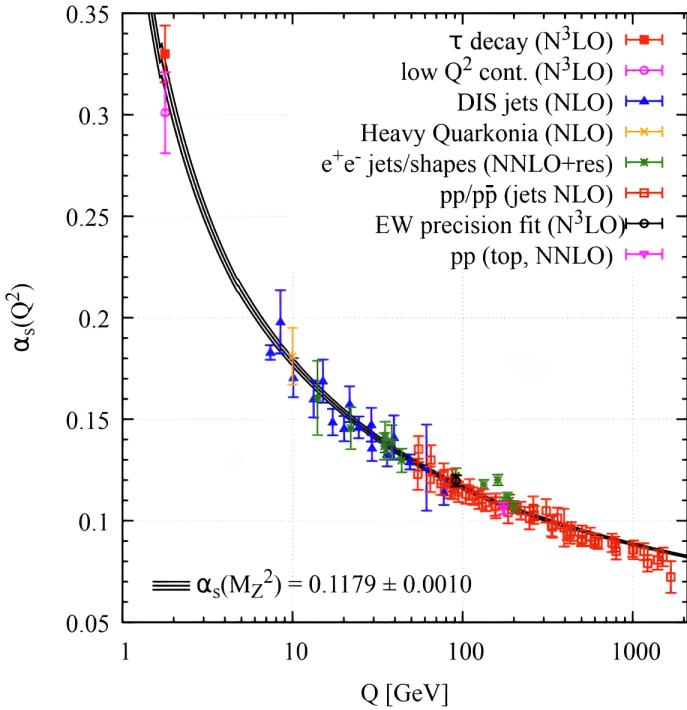


Figure 1.2: Summary of the measurements of α_s as a function of the energy scale Q . The QCD perturbation theory used to extract the value is listed inside the parenthesis. Figure from [1].

1.2.1 Asymptotic freedom

In analogy with QED, the strong coupling constant depends on the momentum transfer Q^2 [52, 53]. It is usually referred to *running coupling constant* and can be described as:

$$\alpha_s(Q^2) = \frac{g_s^2}{4\pi} = \frac{16\pi^2}{(11 - \frac{2}{3}N_f) \ln Q^2/\Lambda_{\text{QCD}}^2}, \quad (1.5)$$

where Λ_{QCD} is the scale parameter and N_f is the number of light quarks. This expression describes how the strength of the interaction between quarks and gluons changes when the energy scale increases, or equivalently the length scale decreases. Unlike the QED effective coupling constant which increases with the energy scale, the α_s decreases with increasing Q^2 . At very high energy scales, the quarks and gluon become *asymptotic freedom* [52, 53], as shown in Fig. 1.2 and confirmed by experimental results. To be specific, for large Q^2 (decreasing distance), α_s decreases, implying that the hadron constituents can be considered as free particles. On the contrary, for decreasing Q^2 (increasing distance), α_s diverges, indicating that the strong interaction is very high, and quarks are strongly bound in hadrons, unable to be separated.

Due to asymptotic freedom, QCD calculations can be treated with different methods at different scales. Perturbative QCD (pQCD) is applicable in high momentum transfer or short-distance scenarios, where



$\alpha_s \ll 1$ decreases logarithmically (very slowly) at the parton level. Therefore, some physics quantities, like cross section, can be computed as a truncated series with varying α_s dependence (α_s^{n+2}). The value of the power n plays a crucial role, with $n = 0$ contributing the most and referred to as *Leading Order (LO)*. As n increases, the contribution diminishes, resulting in more complex diagrams with increased difficulty in calculation. These corrections are termed *Next-to-Leading order (NLO)* for $n = 1$, and *Next-to-Next-to-Leading order (NNLO)* for $n = 2$, and so on. In contrast, for smaller values of Q , α_s gets too large, rendering a perturbative approach inapplicable. QCD turns into a strong-coupling gauge theory, where $\alpha((0.5 \text{ GeV})^2) \sim 1$, indicating the breakdown of perturbation theory validity. In the case of $Q \lesssim 1 \text{ GeV}$, the Green's function of the QCD Lagrangian can still be evaluated on a discrete space-time grid, known as lattice QCD (lQCD) [54]. Notably, there are two outstanding non-perturbative features in the infrared region of QCD: *colour confinement* and the *chiral symmetry breaking* mechanism.

1.2.2 Colour confinement

Colour confinement [55] postulates that charged particles, (anti)quarks and gluons, cannot be isolated. As a consequence, they cannot be directly observed, instead, they always form composite systems known as hadrons. The resulting hadrons are always colour neutral and can be detected. They are classified into two groups: mesons consist of two quarks (one quark and one antiquark), and baryons are composed of three quarks or antiquarks. Because gluons also carry colour, they are affected by colour confinement as well. The confinement of the gluons restricts the range of the strong interaction to a few femtometers. This confinement mechanism is a distinctive characteristic of the strong force in the QCD.

1.2.3 Chiral symmetry breaking

Chiral symmetry breaking [56] generally refers to the spontaneous breaking of chiral symmetry associated with massless fermions. When the mass term is neglected, the QCD Lagrangian defined in Eq. 1.2 exhibits chiral symmetry, implying that the fundamental left- and right-handed parts of the quark fields transform independently. This symmetry of QCD is broken in two ways: explicit breaking representing the theory itself is not invariant under transformations, and spontaneous breaking meaning the chiral symmetry of the Lagrangian is not realized by the ground state. The spontaneous breaking of the chiral symmetry in QCD vacuum [57], results in the appearance of massless Nambu-Goldstone bosons. In reality, pions are much lighter than other hadrons, but they are not massless, which is explained by the mass term for quarks $-m\bar{q}q$ in Lagrangian of QCD. This mass term explicitly mixes the left- and right-handed components of the quarks. At low energy scales, the spontaneous breaking of chiral symmetry is responsible for most of the mass of the hadrons. The explicit breaking of the chiral symmetry contributes a small amount, considering the mass of quarks [58].



1.3 Open heavy-flavour physics

Open heavy-flavour hadrons, particles containing at least one charm (c) or beauty (b) quark with other lighter quarks, are among the most important tools to study QCD in high-energy hadronic collisions. Heavy quarks are predominantly produced in hard-scattering processes in the early stage of collisions and not in later thermal processes. Thus, their primary partonic scatterings and production kinematics are not influenced by the medium effects. Subsequently, they experience the full system evolution, going through the medium expansion and strongly interacting with the medium constituents through inelastic (gluon radiation) and elastic (or collisional) processes. For these reasons, heavy-flavour hadrons offer insights into all stages of the collision evolution. Measurements of heavy flavour production in Pb–Pb collisions allow us to study QGP properties via medium-induced effects. Studies in p–Pb collisions help to disentangle between the final-state and initial-state effects, thereby testing the interplay between soft and hard processes. Moreover, measurements in pp collisions serve as crucial reference for p–Pb and Pb–Pb measurements, as well as essential tests for calculations based on pQCD down to low transverse momentum. Therefore, a comprehensive understanding of heavy-quark production in pp collisions is important, and is the main focus of this thesis.

In this chapter, a brief introduction of the open heavy-flavour production in pp collisions is presented in Section 1.3.1. The current status of the charm hadron studies is reported, followed by discussions about the heavy quark hadronisation mechanism shown in Section 1.3.2, particularly focusing on the physics behind the two analyses in this thesis.

1.3.1 Open heavy-flavour production in pp collisions

In pp collisions, protons consist of quarks and gluons referred to as *partons*. There are one or more inelastic parton-parton interactions involved in inelastic pp collisions. It offers a good opportunity to probe the pQCD predictions, given that the larger Q^2 required for the $c\bar{c}$ and $b\bar{b}$ production implies a α_s significantly lower than unity, where the production cross section of $q\bar{q}$ can be calculated with perturbative calculations.

The strong interaction, conserving flavour charge, requires heavy quarks Q to be produced in quark-antiquark pairs. The main contributing processes at leading-order (LO), $\mathcal{O}(\alpha_s^2)$, are the quark-antiquark annihilation $q\bar{q} \rightarrow Q\bar{Q}$, and the gluon fusion $gg \rightarrow Q\bar{Q}$ [2], which is shown in Fig. 1.3. At next-to-leading order (NLO), corresponding to $\mathcal{O}(\alpha_s^3)$, flavour-excitation and gluon-splitting processes also contribute to heavy-quark hadron production as well, presented by $gq \rightarrow Q\bar{Q}q$ or $gg \rightarrow Q\bar{Q}g$, as seen in Fig. 1.4. The minimum virtuality of these scatterings is $Q^2 > 4M_{c,b}^2$, restricted to the hard scatterings



in the initial stages of the collisions.

In pQCD calculations, the production cross section of heavy-flavour hadrons (H_Q) $d\sigma_{pp \rightarrow H_Q + X}$, is usually calculated using the factorisation approach as a convolution of three factors [59]: i) the Parton Distribution Functions (PDFs); ii) the hard-scattering cross section at the partonic level; and iii) the fragmentation functions of the produced heavy quarks into given species of heavy-flavour hadrons. It can be written as in Eq. 1.6:

$$\begin{aligned} d\sigma_{pp \rightarrow H_Q + X}(\sqrt{s}) &= \sum_{i,j=q,\bar{q},g} f_1(x_i, \mu_F^2) f_2(x_j, \mu_F^2) \\ &\otimes d\sigma_{ij \rightarrow Q\bar{Q}}^{\text{hard}}(\alpha_s(\mu_R^2), \mu_F^2, M_Q, x_i x_j s) \\ &\otimes D_Q^{H_Q}(z, \mu_F^2), \end{aligned} \quad (1.6)$$

where considering two colliding protons, labeled as ‘1’ and ‘2’. Q is the heavy quark (charm or beauty), M_Q is its mass. The sum runs over all possible sub-process leading to the heavy-flavour hadron.

- $f_i(x_i, \mu_F^2)$ is the *parton distribution functions*, which represents the probability of finding a quark or a gluon with a momentum fraction of the original proton $x_i = p_{\text{parton}}/p_p$, called *Bjorken-x*. The PDFs are evolved with the virtuality Q^2 up to the factorisation scale μ_F , at which the parton density is evaluated.
- $d\sigma_{ij \rightarrow Q\bar{Q}}^{\text{hard}}$ is the partonic cross-section of the $Q\bar{Q}$ production process. Given that the large mass of the quarks involved ($M_b > M_c > \Lambda_{QCD}$), it is related to interactions of partons at high Q^2 , where it is computed by the pQCD. Besides the μ_F , The partonic cross section also depends on the renormalisation scale μ_R at which the strong coupling constant is evaluated.
- $D_Q^{H_Q}(z, \mu_F^2)$ is the *fragmentation function (FF)*, denoting the probability for a given heavy quark Q hadronise into a specific hadron H_Q carrying a momentum fraction $z = p_{H_Q}/p_Q$, called *fragmentation fraction*. Heavy quarks and anti-quarks are separately fragmented into hadrons for heavy

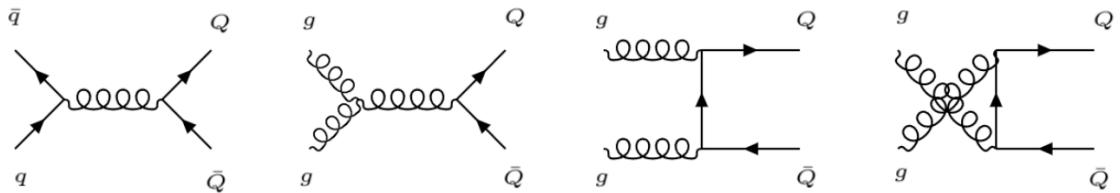


Figure 1.3: Feynman diagram of leading order quark-antiquark annihilation and gluon fusion for the production of $c\bar{c}$ and $b\bar{b}$. Figure from [2].

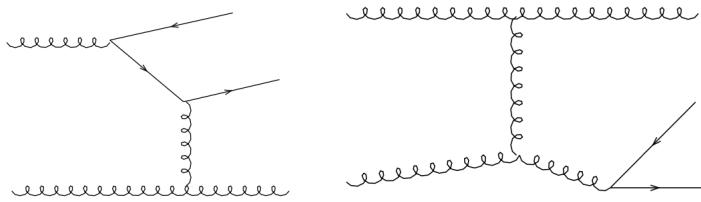


Figure 1.4: Feynman diagram of next-to-leading order, left panel: flavour-excitation, right panel: gluon-splitting. Figure from [2].

flavour production.

Both the PDFs and FFs describe non-perturbative processes and cannot be obtained with pQCD calculations. The PDFs are typically parameterised using measurements of deep-inelastic scattering [60–62], while the FFs are usually parameterised by measurements in e^+e^- collisions [63], assuming universality among different collision systems. However, recent measurements are questioning this assumption, more discussion will be shown later in this chapter. Both the PDFs and FFs are parameterised at a given energy scale and subsequently are evolved to the desired scale by using the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [64–66]. The two scales, μ_R , μ_F , are usually chosen of the same order of the momentum transfer of the hard process $\mu_R \sim \mu_F \sim \sqrt{M_Q^2 + p_{T,Q}^2}$ or $\sqrt{4M_Q^2 + p_{T,Q}^2}$.

At present, state-of-the-art perturbative calculations that provide predictions for the production cross sections of the open heavy flavours, exploiting the collinear factorisation approach, are performed with the Fixed-Order-Next-to-Leading-Log (FONLL) [3, 4] and General Mass Variable Flavour Number Scheme (GM-VFNS) [5, 6] frameworks. In Fig. 1.5, both predictions are compared to the p_T -differential production cross section of D^0 , D^+ , and B^0 mesons measured in pp collisions at $\sqrt{s} = 5.02$ TeV and $\sqrt{s} = 7$ TeV by the ALICE and LHCb Collaborations, respectively. The pQCD calculations are compatible with the measured beauty and charm production at the LHC within their theoretical uncertainties. However, the measurements of the charm mesons systematically lie at the upper edge of FONLL uncertainty band, which comes from the variations of μ_F , μ_R , M_Q , and the associated uncertainty of the PDFs.

In addition to the collinear-factorisation approach for calculating the charm cross-section production, an effective and efficient alternative is the so-called k_T -factorisation approach [67–69]. It performs the calculations at LO taking into account the transverse momentum (k_T) of both partons involved in the hard process. In the bottom left panel of Fig. 1.5, the comparison between k_T -factorisation and the measured production of D^0 mesons can be found. There is good agreement at low and intermediate p_T ; however, k_T -factorisation predicts higher values for $p_T > 7$ GeV/c.

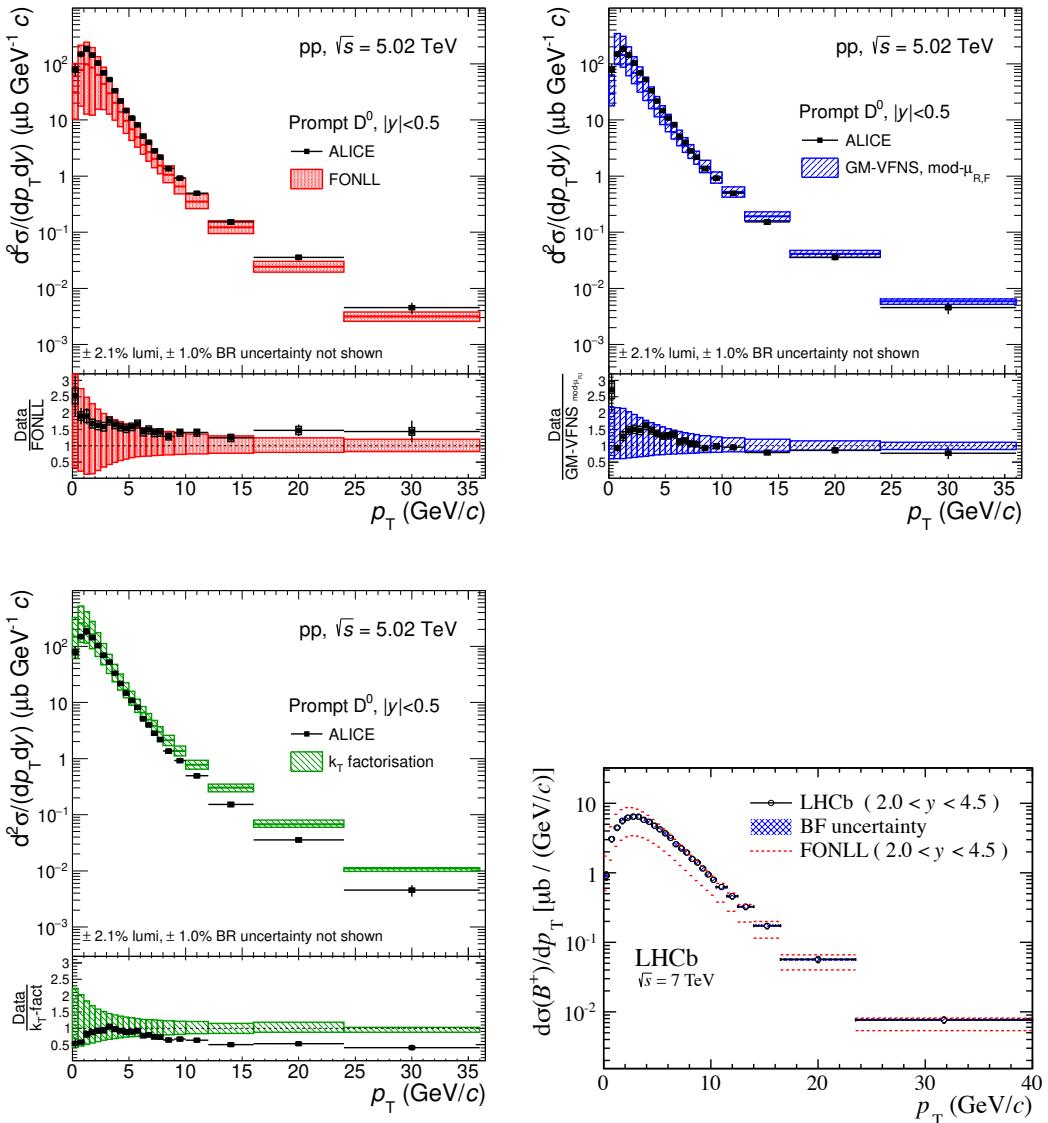


Figure 1.5: Top two panels and bottom left panel: p_T -differential production cross section of D^0 mesons at mid-rapidity ($|y| < 0.5$) measured by the ALICE Collaboration in pp collisions at $\sqrt{s} = 5.02 \text{ TeV}$ compared to FONLL [3, 4], GM-VFNS (mod- $\mu_{R,F}$) [5, 6], and k_T -factorisation [7]. Figures from [8]. Bottom right panel: p_T -differential production cross section of B^+ mesons at forward rapidity ($2.0 < y < 4.5$) measured by the LHCb Collaboration in pp collisions at $\sqrt{s} = 7 \text{ TeV}$ compared to FONLL calculations. Figure from [9].

Another approach to describe the production of heavy-flavour hadrons, an alternative to pQCD calculations, is the general-purpose Monte-Carlo (MC) generators, like PYTHIA [13] and HERWIG [70]. These generators give a more complete description of the final state, including hard and soft interactions, the initial and final-state parton showers, hadronisation processes, multiple parton interactions, fragmentation, and decays. However, their disadvantage is that they only implement LO processes, while higher order calculations are included only in an approximate approach. Monte Carlo genera-

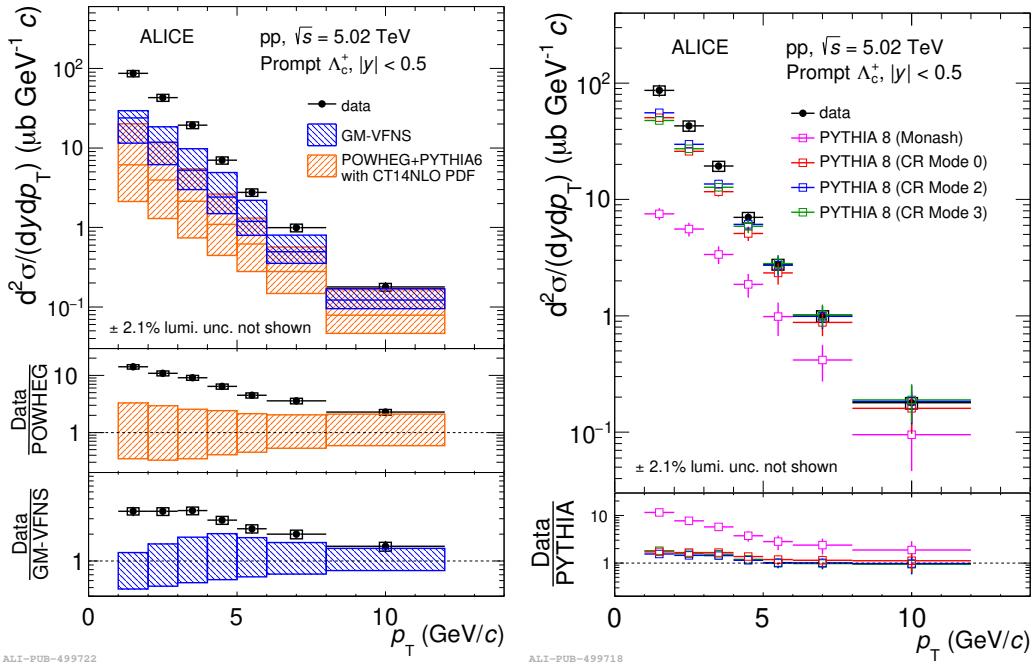


Figure 1.6: The p_T -differential production cross section for prompt Λ_c^+ in $\sqrt{s} = 5.02$ TeV in pp collisions at ALICE Collaboration [10, 11]. Comparisons to predictions from GM-VFNS [5, 6], POWHEG [12], and PYTHIA 8 [13] are shown, which all significantly underestimate the data at low p_T .

tors including the NLO processes in the hard scattering, like POWHEG [12] and MC@NLO [71], are available as well. They can be compared to the parton shower of PYTHIA and HERWIG for a compete modelling. Comparisons of POWHEG + PYTHIA calculations to the measurements of D-meson spectra and to the collinear factorisation calculations can be found in Ref [72]. The predictions agree within the theoretical uncertainty band, dominated by the variations of the two scales (μ_R and μ_F), though the central values give a lower trend than the data.

It is important to note that all statements on data-theory predictions mentioned above are intentionally made for heavy-flavour mesons. Recent measurements of heavy-flavour baryon production at the LHC [10, 11, 21, 73–77] show deviation with those theoretical calculations, which underestimate the data. Fig. 1.6 gives an example of the prompt Λ_c^+ -baryon production cross section in pp collisions at $\sqrt{s} = 5.02$ TeV in ALICE Collaboration [10, 11]. The predictions from GM-VFNS [5, 6], POWHEG [12], and PYTHIA 8 Monash [13] all significantly underestimate the data at low p_T .

Attention has been drawn in recent years, as shown in the right panel of Fig. 1.6, where three new predictions match better the data. Further details regarding this puzzle and its explanation will be discussed in Section 1.3.2.



1.3.2 Charm quark hadronisation mechanisms

The clean environment in e^+e^- collisions presents many advantages to study fragmentation processes of heavy quarks with respect to hadronic collisions. In e^+e^- collisions, there is no hadronic initial state, therefore, concurrent partonic interactions and beam remnants are not expected. The precise theoretical knowledge of the total $c\bar{c}$ cross section gives the chance to directly measure the fragmentation fraction (FF) of charm quarks to a given charm-hadron species, $f(c \rightarrow H_c)$, without the need of measuring the cross section of the other charm-hadron species. At LEP, it was confirmed that the total charm-production cross section is consistent with the SM expectation, with the sum of the measured D^0 , D^+ , D_s^+ and Λ_c^+ , and of an assumed 2% additional contribution from $\Xi_c^{+,0}$ and Ω_c^0 . The charm fragmentation fractions $f(c \rightarrow h_c)$ have been measured in e^+e^- and e^-p collisions by different experiments [14], as shown in Fig. 1.7. The charm-hadron production measurements in e^+e^- collisions at B factories [78–85] and in Z^0 -boson decays [86–90], were studied together with data from photoproduction (PHP) and deep-inelastic scattering (DIS) in e^-p collisions at HEAR [91–95], and also with the first data from pp collisions at the LHC. The comparison shows that there is no significant difference observed among the charm FF measured in the various collision systems and collision energies. The measurements of FF in e^+e^- and e^-p are compatible within uncertainties, on which the fragmentation universality assumption was built. It was also expected to hold in pp collisions at the LHC. However, as discussed in the next, it turns out not to be true.

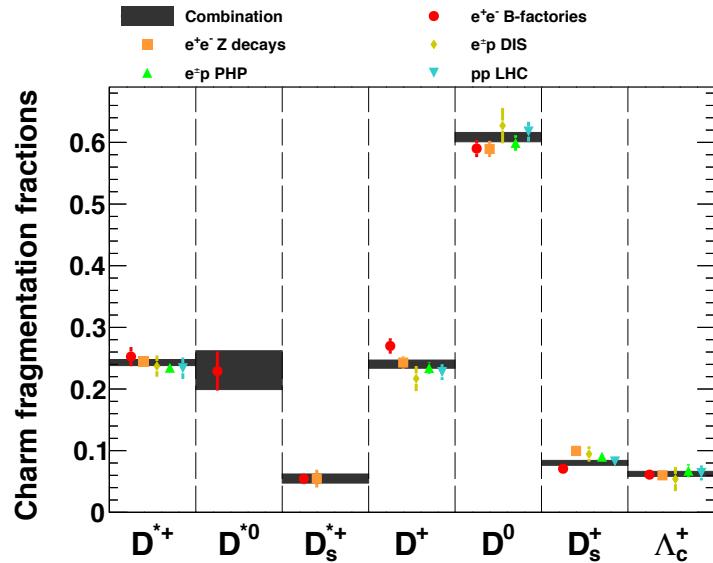


Figure 1.7: The charm fragmentation fractions $f(c \rightarrow h_c)$ as measured in e^+e^- and e^-p collisions by different experiments [14]. Fractions for pp collisions are shown as well, but these are directly based on the Λ_c^+ production measurement of LHCb [15].

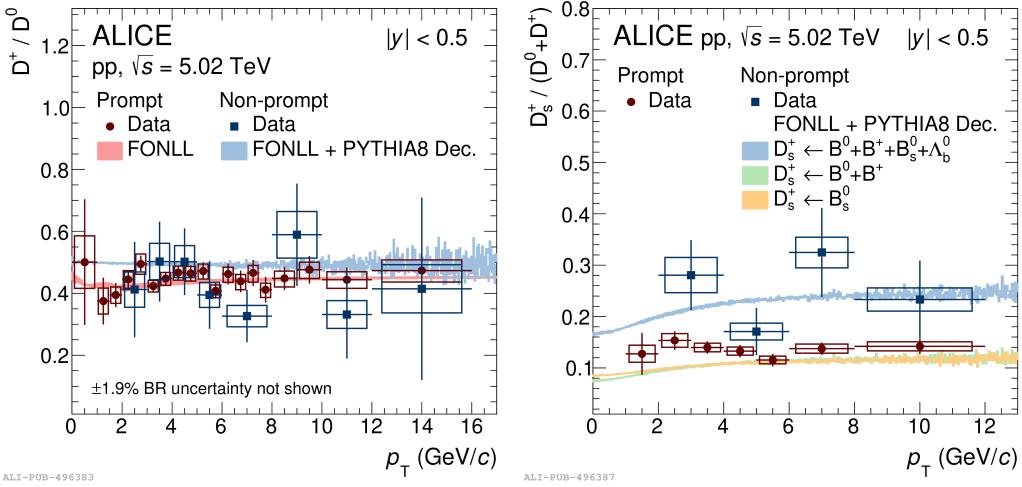


Figure 1.8: Left panel: ratios between the p_T -differential production cross section of D^+ and D^0 mesons [16]. Right panel: ratios between the D_s^+ -meson and the sum of the D^0 - and D^+ -meson production cross sections. The ratios are compared with prediction obtained with FONLL calculations [3, 4] using PYTHIA 8 [17, 18] for the $H_b \rightarrow D + X$ decay kinematics. For the non-prompt $D_s^+/(D^0 + D^+)$ ratio, the predictions for the D_s^+ from B_s^0 and from non-strange B-meson decays are also displayed separately.

Recently, the ALICE Collaboration reported comprehensive measurements of the ground-state charm hadrons, the charm mesons D^0 , D^+ , D_s^+ , D^{*+} [8, 16] and the charm baryons Λ_c^+ , $\Xi_c^{+,0}$, Ω_c^0 [19, 21, 25, 33]. A lot of interest in the field has focused on studying heavy-quark hadronisation, particularly via the measurement of hadron-to-hadron yield ratios, because the PDFs and partonic interaction cross section are common to the charm- and beauty hadron species and, therefore, cancel out in the yield ratios.

The meson-to-meson yield ratios, both for the respective prompt (i.e. produced in the hadronisation of charm quarks or from the decay of excited open charm and charmonium states) and non-prompt (coming from beauty-hadron decays) components of the D-meson production, are well described by pQCD calculations using fragmentation functions tuned on e^+e^- and e^-p collision measurements [16, 96, 97], as illustrated in Fig. 1.8. The ratios are also well described by the Lund string fragmentation model [98], which is a key model of hadronization, implemented in the PYTHIA 8 event generator. The PYTHIA 8 event generator with the Monash 2013 tune, where two partons are connected by colour string fragments into hadrons, is adopted in the leading-colour approximation. This configuration suppresses the connection of quarks and gluons originating from independent parton scatterings, resulting in a similar picture of heavy-quark fragmentation and hadronisation in hadronic collisions compared to those in e^+e^- collisions. Unfortunately, these calculations underestimate recent measurements of charm baryon-to-meson production ratios in hadronic collisions [19, 22, 99] and fragmentation fractions in pp collisions, demonstrating that the assumption of universality of the hadronisation process across collision systems has to be reconsidered.

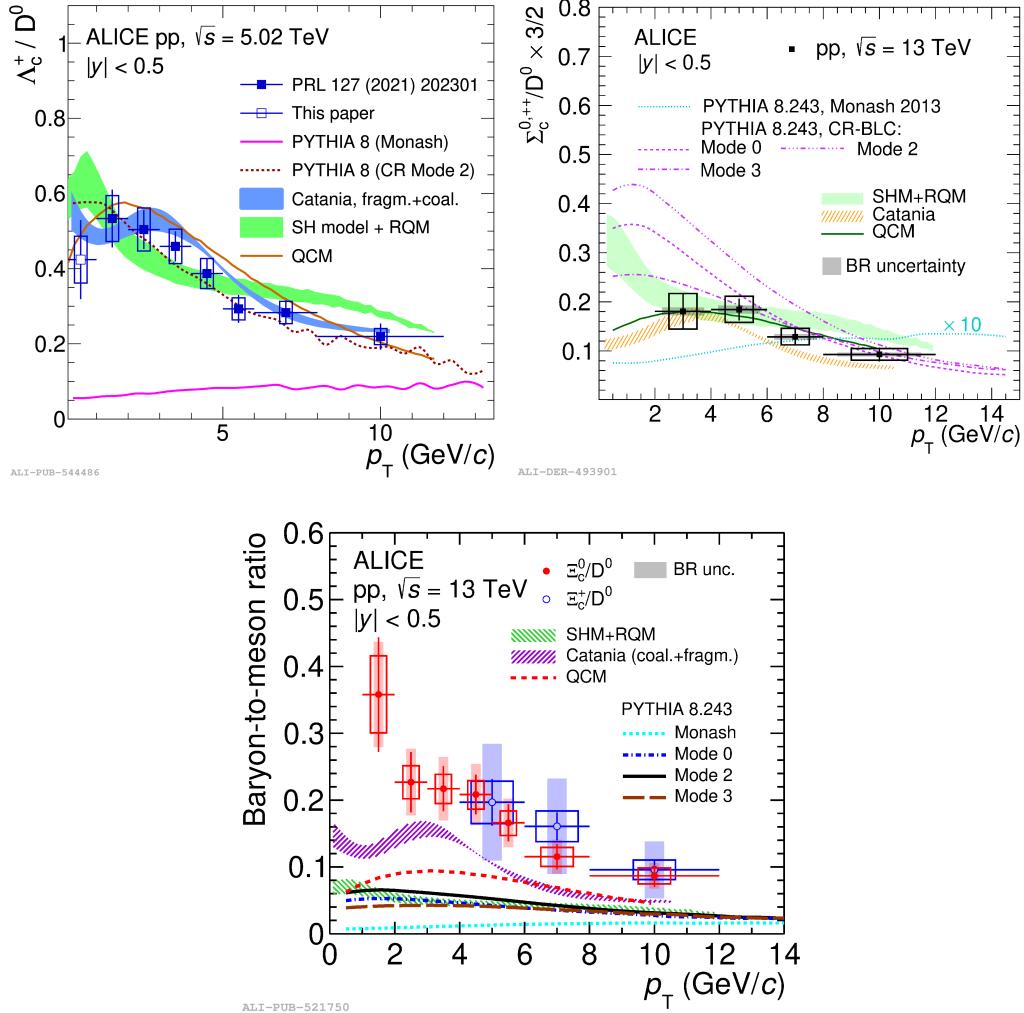


Figure 1.9: Top left panel: the Λ_c^+/D^0 ratios a function of p_T measured in pp collisions at $\sqrt{s} = 5.02$ TeV [19]. Top right panel: Prompt $\Sigma_c^{0,++}/D^0$ yield ratio in pp collisions at $\sqrt{s} = 13$ TeV [20]. Bottom panel: the $\Xi_c^{0,+}/D^0$ ratios a function of p_T measured in pp collisions at $\sqrt{s} = 13$ TeV [21].

The Λ_c^+/D^0 baryon-to-meson yield ratio as a function of p_T at midrapidity in pp collisions at $\sqrt{s} = 5.02$ TeV is shown in the top left panel of Fig. 1.9. The ratio value is about 0.4 - 0.5 at low p_T region, while decreasing to be around 0.2 at high p_T . There is a strong p_T -dependence of the Λ_c^+/D^0 ratio, compared with the ratios shown in Fig. 1.8 without a significant p_T dependence, therefore, indicating that the fragmentation functions of baryons and mesons differ significantly. The Λ_c^+/D^0 ratios are compared with model calculations implementing different hadronisation processes. This first measurement of Λ_c^+ down to $p_T = 0$ probes a very important range to investigate hadronisation, as a substantial fraction of the total charm cross section is contained at $p_T < 1$ GeV/c. At low p_T , the ratio is much larger than the PYTHIA 8 Monash predicted by string fragmentation models measured in e^+e^- data. However, the



ratios are qualitatively described by modified hadronisation processes, which will be explained in the following.

A similar enhancement with respect to e^+e^- collisions is observed for the $\Sigma_c^{0,++}/D^0$ yield ratios shown in the top right panel of Fig. 1.9, indicating that the enhancement of Λ_c^+/D^0 can be partially explained by the $\Sigma_c^{0,++}$ feed-down. The $\Sigma_c^{0,++}/D^0$ ratio is close to 0.2 for $2 < p_T < 6 \text{ GeV}/c$, and about 0.1 for $8 < p_T < 12 \text{ GeV}/c$, though the uncertainties do not allow to draw conclusions about the p_T dependence. From the measurements in BELLE Collaboration (Table IV in Ref. [100]), the ratio of $\Sigma_c^{0,++}/\Lambda_c^+$ was measured to be around 0.17 in e^+e^- collisions at $\sqrt{s} = 10.52 \text{ GeV}$, thus, the $\Sigma_c^{0,++}/D^0$ ratio can be determined to be around 0.02. Additional information shows the remarkable difference between pp and e^+e^- collision systems.

In the bottom panel of Fig. 1.9, the measurement of charm baryons is extended to those containing strange quarks, production of $\Xi_c^{+,0}$ measured in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ [21]. The two ratios show similar trend of baryon-to-meson as what was observed for the Λ_c^+/D^0 ratio, while the $\Xi_c^{+,0}/D^0$ ratios are generally lower. Those models qualitatively capture well for the Λ_c^+/D^0 ratio as shown in the top left panel of Fig. 1.9, however, significantly predict low values for the charm-strange baryon states $\Xi_c^{+,0}$. It has been recently also argued that an enhancement of $\Xi_c^{+,0}/D^0$ could be due to the large attraction of the diquark (us and ds) channel with respect to the ud one. Another interesting observation is given by the fact that the baryon-to-baryon yield ratio $\Xi_c^{+,0}/\Sigma_c^{0,++}$ in Ref. [21] is described well by the PYTHIA 8 Monash tune, in contrast, which largely underestimate both $\Xi_c^{+,0}/D^0$ and $\Sigma_c^{0,++}/D^0$ ratios. This may suggest that the production of the two baryons could be both suppressed in e^+e^- collisions due to similar mechanisms.

With all the available measurements of hadrons in ALICE, the charm-quark fragmentation fractions at midrapidity in pp collisions at $\sqrt{s} = 5.02 \text{ TeV}$ and 13 TeV are shown in Fig. 1.10, indicating no significant energy dependence within the current uncertainties. In addition, a clear difference is observed with respect to the charm fragmentation fractions measured in e^+e^- and e^-p collisions. Since the fragmentation fractions sum to unity, there is a reduction of the relative D-meson abundance by an about factor 1.5 relative to e^+e^- and e^-p , taking into account the enhancement of baryon production. These results confirm that the baryon enhancement at the LHC with respect to e^+e^- and e^-p collisions, is the effect of different hadronisation mechanisms happening in the parton-rich environment produced in pp collisions, independent of the centre-of-mass energy.

To explain this phenomenon of the baryon enhancement, several models have been proposed recently

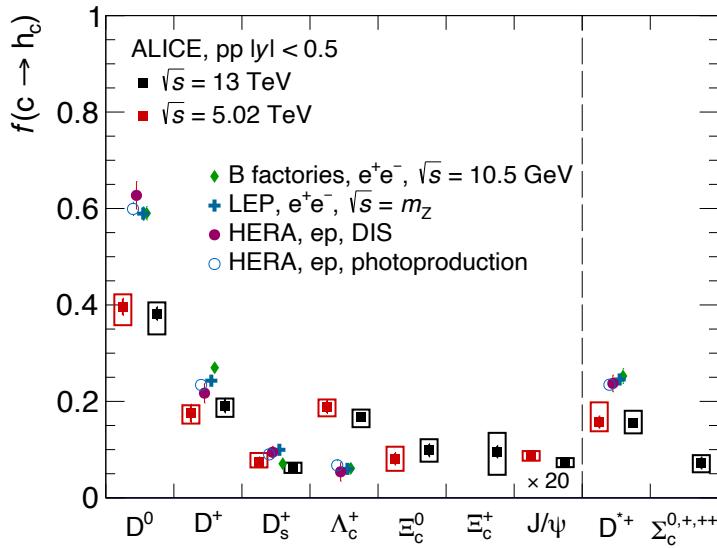


Figure 1.10: The charm fragmentation fractions at midrapidity ($|y| < 0.5$) in pp collisions at $\sqrt{s} = 5.02$ TeV and $\sqrt{s} = 13$ TeV compared with results in e^+e^- and e^-p collisions [14, 22]. The fragmentation fractions $f(c \rightarrow h_c)$ of J/ψ mesons are multiplied by a factor 20 for better visibility.

based on different developments and assumptions: 1) include dynamical processes that are relevant in quark-and-gluon enriched systems, colour reconnection beyond leading colour approximation, quark recombination, and Catania; 2) treat hadronisation as a statistical process, considering a set of yet-unobserved higher-mass charm-baryon states. In the following, a brief introduction to these models is presented.

1.3.2.1 PYTHIA CR Mode 0, 2, 3

Traditionally, MC generators make use of the *leading-colour (LC)* approximation, which tracks the colour flow on an event-by-event basis. This approach assumes that each quark is colour-connected to a unique colour parton in the event, similar to a leading-colour QCD dipole, where gluons carry both a colour and an anticolour charge, being then connected to two other partons. In terms of string, it can be interpreted as gluons forming transverse “kinks” on strings, whose endpoints are quarks and antiquarks, as dictated by the Lund model. This approach provides a good approximation of the hadronisation in e^+e^- collisions, however, is challenged when moving to hadronic collision systems, such as pp collisions. In this case, the initial state is sensitive to the strong force, but the partons inside protons carry colour charge and coloured beam remnants. In pp collisions, *multiple parton interactions (MPI)* have dominated contribution, and they are crucial to explain the underlying event and soft-inclusive (minimum-bias/pileup) physics, especially at high particle multiplicities. In such a collision, the possibility of phase-space overlaps between the particles produced in different MPI events is non trivial.



Earlier works [101, 102] already took into account the possibility for the different MPI systems to be colour-connected within a colour “chain” to minimise the total colour charge. However, the effects of colour reconnection were estimated to be small. In particular, this approach estimates the average p_T of charged particles in such events to be flat as a function of particle multiplicities. But the experimental results exclude this possibility, remarking on the urgent need to introduce non-trivial colour reconnection effects [103–105].

The pure SU(3) group structure of QCD provides some sub-leading combinations among different partons that may create colour singlet before the single parton fragmentation. However, these additional combinations are insufficient to describe the measured baryon production in hadronic collisions. One model important for describing the baryon production in pp collisions implementing PYTHIA is introduced, which is the so-called QCD-based colour reconnection (CR) model [23]. Extending the string model beyond the LC approximation reintroduces colour-space ambiguities, considering the $N_c = 3$ finite colour structure of QCD, namely, *colour reconnection beyond leading colour (CR-BLC)* approximation. It assumes that the main subleading parton-parton combination possibilities of real QCD are encoded by 9 colour indices, leading to 9 possible states of quarks (anti-quarks), and 72 states for gluons. In the other word, modelling partons using SU(3) colour-space means that each colour is not necessarily unique, allowing multiple different partons to be colour-connected and leading ambiguity in where confining potentials rise.

To determine between which partons the confining fields form, the configuration that minimises the string “length” or the action potential in the systems, has to be made. This “string length” is a Lorentz-invariant measure, denoted as λ *string-length measure*, which is interpreted as a measure of the energy density per unit length of the string. To be specific, it is the area spanned by the string prior to hadronisation, namely the potential energy of the string. By requiring the minimised string lengths lower than the λ measure between the colour-connected partons, the strings are thus able to form between partons instead of being produced in the same hard scattering, allowing topologies to span across different MPIs. In this new framework, all the possible combinations associated with overlaps in colour-space are considered. These overlap become relevant for hadronic collisions at high energies, particularly where MPIs produce densely populated partonic environments. These additional combinations are the so-called *junctions*, reported in the right panel of Fig. 1.11.

Considering the λ -measure for junctions brings additional approximation, several tuneable parameters are added to control the junction procedure:

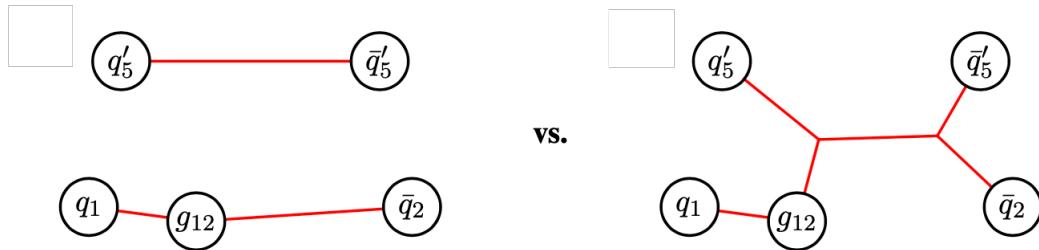


Figure 1.11: Schematic of a multi-parton state with index assignments resulting in a junction-type colour-space ambiguity. Left: The original (LC) string topology. Right: An alternative string topology with a junction and an antijunction, allowed by cyclically matching ‘2’ and ‘5’ indices. Figure from [23].

- $C_j \equiv m_{0j}/m_0$: a higher m_{0j} means a lower λ measure and a value of C_j above unity, represents an enhancement in junction production, and vice versa.
- C_{time} : the reconnections between two string pieces should happen causally. If either string hadronised before the other forms, results in no space-time region where reconnections could occur physically. The time when strings begin to hadronise is related to the inverse of Λ_{QCD} , $\tau_{\text{had}} \sim 1/\Lambda_{QCD}$. In the rest frame of a hadronising string piece, the formation time of the QCD dipole accordingly can be given roughly by the inverse of its invariant mass, $\tau_{\text{form}} \sim 1/m_{\text{string}}$. Therefore, the two strings should be resolved each other during the time between the formation and hadronisation, by considering the time dilation occurring due to the relative boost. The two strings can be defined to be in causal contact if the relative boost parameter meets, Eq. 1.7:

$$\gamma \tau_{\text{form}} < C_{\text{time}} \tau_{\text{had}} \Rightarrow \frac{\gamma c}{m_{\text{string}} r_{\text{had}}} < C_{\text{time}}, \quad (1.7)$$

where C_{time} is tuneable parameter and $r_{\text{had}} (= r_{\text{had}}c) \equiv 1\text{fm}$ is a fixed constant given by the typical hadronisation scale.

The different modes of this CR model depend on the causal connection requirement. Several options are available for the choice of CR time-dilation methods. There are three modes as follows, used recently to compare with the data, as shown in Fig. 1.9.

- Mode 0: no time-dilation constraints. m_0 controls the amount of CR
- Mode 2: the causal contact is required to be valid among all the dipoles involved in a reconnection, which is the strict condition
- Mode 3: the time dilation as in Mode 2, but requiring only a single connection to be causally connected, which is loose case



The parameters of the models are tuned by taking experimental results, the junction topologies significantly enhance the production of open charmed baryons in pp collisions at the LHC energies. Fig. 1.9 shows that this model describes the Λ_c^+/D^0 and $\Sigma_c^{0,++}/D^0$ ratios, but underestimates the $\Xi_c^{0,+}/D^0$ ratio. It is still a puzzle of failing to describe the heavier charm-strange baryon, motivating an effort from the experimental side to provide more precise measurements of their production in pp collisions at the LHC.

1.3.2.2 Quark (re-)combination mechanism

In Quark (re-)combination mechanism (QCM) model [36], the charm quark c is combined with a co-moving (equal-velocity) light antiquark or two co-moving quarks to form a charm meson or baryon. The momentum of charm mesons and baryons is characterised $p_H = p_c + p_{\bar{q},qq}$. This (re)-combination mechanism of charm quark hadronisation will effect the momentum spectrum of charm hadrons, straightforwardly, influencing the yield ratios of charm baryon to charm meson.

The momentum distribution of a single-charm meson $M_{c\bar{l}}$ and baryon $B_{cll'}$, (l and l' indicating two light-flavour quarks among u, d, s), can be interplayed as:

$$f_{M_{c\bar{l}}}(p) = \int dp_1 dp_2 f_{c\bar{l}}(p_1, p_2) \mathcal{R}_{M_{c\bar{l}}}(p_1, p_2; p), \quad (1.8)$$

$$f_{B_{cll'}}(p) = \int dp_1 dp_2 dp_3 f_{cll'}(p_1, p_2, p_3) \mathcal{R}_{B_{cll'}}(p_1, p_2, p_3; p), \quad (1.9)$$

where $f_{ci}(p_1, p_2)$ is the joint momentum distribution for charm quark (c) and light anti-quark (\bar{l}). The $\mathcal{R}_{M_{c\bar{l}}}(p_1, p_2; p)$ is the probability density for the c and \bar{l} with momentum p_1, p_2 , to be combined into a meson $M_{c\bar{l}}$, with the momentum p . Similar explanation for a baryon $B_{cll'}$,

The momentum integrated yield for charm meson and baryon is as follows, respectively:

$$N_{M_{c\bar{l}}} = N_c N_{\bar{l}} \frac{\kappa_{M_{c\bar{l}}}}{A_{M_{c\bar{l}}}} = N_c N_{\bar{l}} \mathcal{R}_{c\bar{l} \rightarrow M_{c\bar{l}}}, \quad (1.10)$$

$$N_{B_{cll'}} = N_c N_l N_{l'} \frac{\kappa_{B_{cll'}}}{A_{B_{cll'}}} = N_c N_l N_{l'} \mathcal{R}_{cll' \rightarrow B_{cll'}}, \quad (1.11)$$

where $\mathcal{R}_{c\bar{l} \rightarrow M_{c\bar{l}}}$ and $\mathcal{R}_{cll' \rightarrow B_{cll'}}$ corresponds to the momentum-integrated probability for c and light-flavour quark(s) to produce a meson $M_{c\bar{l}}$ and a baryon $B_{cll'}$.

The p_T distributions of the quarks at hadronisation are the inputs for this model. Given these inputs, the QCM model describes well for Λ_c^+/D^0 and $\Sigma_c^{0,++}/D^0$ ratios shown in Fig. 1.9, both the magnitude



and the p_T dependence within the experimental uncertainties. The QCM gives better predictions than other models not including the charm quark coalescence that significantly underestimates the data. This comparison suggests a new scenario of the low p_T charm quark hadronisation in the presence of the underlying light quark source in pp collisions.

1.3.2.3 Catania model

In 2009, the coalescence plus independent fragmentation approach was proposed already [106]. It predicted a quite large Λ_c production in Pb–Pb collisions, $\Lambda_c/D^0 \sim O(1)$, about one order of magnitude larger than the one measured in e^+e^- collisions where the fraction of baryon production about $f(c \rightarrow \Lambda_c^+ \sim 0.06)$ [14]. Measurements from STAR Collaboration in Au–Au at $\sqrt{s_{NN}} = 200$ GeV [107] confirmed such a prediction. Also recent measurements in pp collisions at the LHC [10, 11, 108–110] show similar phenomenologies as what observed in nucleus-nucleus collisions where these results have been justified by the formation of QGP. These phenomena include strangeness enhancement, enhancement of the baryon-to-meson yield ratios (discussed above), and the ridge and large collectivity, giving hints that there could be a possibility of QGP expected in pp collisions.

Different theoretical studies [111–113] with hydrodynamical and transport-based calculations are able to provide reasonable description of p_T spectra and two-particle correlations, pointing to the possible formation of a hot QCD matter at energy density larger than the pesudo-critical one, with a lifetime $\tau \approx 2 \text{ fm}/c$. Based on this, Catania model approach implements hadronisation of charm quarks via coalescence with independent fragmentation and assumes the presence of a flowing thermalised medium of light quarks (u, d, s). It is characterised with lifetime $\tau = 2.5 \text{ fm}/c$, transverse radius of the fireball $R = 2 \text{ fm}$ and the temperature of the bulk $T_c = 165 \text{ MeV}$, according to hydro-dynamical simulations [111]. In the calculation, the assumption is made, that charm quarks that do not hadronise via coalescence are converted to hadrons by the fragmentation process. Therefore, a fragmentation probability can be given by $P_{frag}(p_T) = 1 - P_{coal}^{total}(p_T)$, where P_{coal}^{total} represents the total probability of given quarks coalesced to hadrons. In this model, the coalescence is evaluated by the *Wigner function*, in which the momentum spectrum of hadrons from the coalescence can be given by, Eq. 1.12:

$$\frac{dN_H^{\text{coal}}}{dyd^2P_T} = g_H \int \prod_{i=1}^{N_q} \frac{d^3p_i}{(2\pi)^3 E_i} p_i \cdot d\sigma_i f_{qi}(x_i, p_i) \\ \times f_H(x_1 \dots x_{N_q}, p_1 \dots p_{N_q}) \delta^{(2)} \left(P_T - \sum_{i=1}^n p_{T,i} \right), \quad (1.12)$$

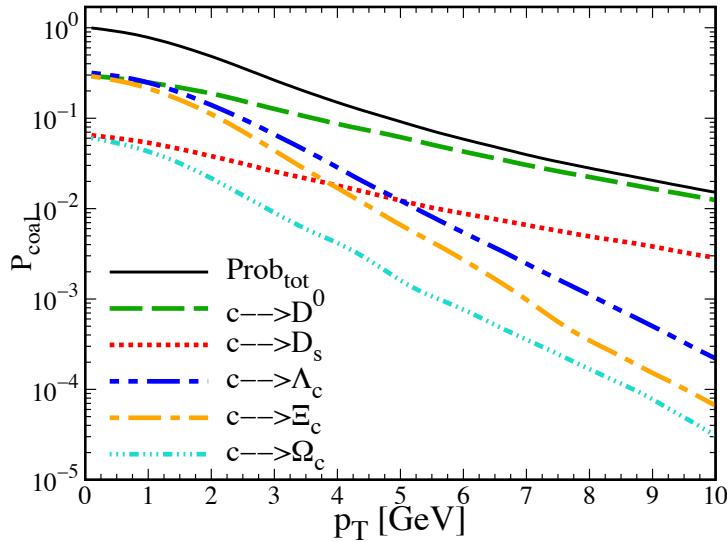


Figure 1.12: The charm quark coalescence probability as a function of the charm quark p_T for pp collisions at LHC. The different coloured lines correspond to the probability of the charm quark to coalesce in a given hadron species, while the black line represents the total coalescence probability. Figure from [24].

where g_H is the statistical degeneracy factor to form a colourless hadrons from quarks and antiquarks with spin 1/2 and, correspondingly $g_D = 1/36$ for mesons and $g_B = 1/108$ for baryons. $d\sigma_i$ is an element of a space-like hypersurface. f_{q_i} are the quarks and anti-quarks phase-space distribution functions for i -th quark or antiquark. The last term $f_H(x_1 \dots x_{N_q}, p_1 \dots p_{N_q})$ represents the Wigner function, telling the probability for a hadron to be produced from quarks with given phase-space coordinates.

The momentum spectrum of hadrons from the charm parton fragmentation can be written as:

$$\frac{dN_H^{\text{fragm}}}{d^2 p_T dy} = \sum \int dz \frac{dN^{\text{fragm}}}{d^2 p_T dy} \frac{D_{H/c}(z, Q^2)}{z^2}, \quad (1.13)$$

where $D_{H/c}(z, Q^2)$ corresponds to the fragmentation function, representing the probability of charm quark to produce a hadron H with a fraction z of the original charm momentum, taking the $Q^2 = (p_{\text{had}}/2z)^2$ the momentum scale of the fragmentation process. Eq. 1.13 is coming from the Peterson fragmentation function [114], commonly used for charm quarks.

The interplay between fragmentation and coalescence depends on the charm p_T . In Fig. 1.12, the coalescence probability of the charm quarks to the various hadron species is shown as a function of the charm p_T . It shows that the P_{coal} is high for low quark momentum and quickly decreases with increasing trans-



$n_i \cdot 10^{-4} \text{ fm}^{-3}$	D^0	D^+	D^{*+}	D_s^+	Λ_c^+	$\Xi_c^{+,0}$	Ω_c^0
PDG(170)	1.161	0.5098	0.5010	0.3165	0.3310	0.0874	0.0064
RQM(170)	1.161	0.5098	0.5010	0.3165	0.6613	0.1173	0.0144

Table 1.1: Thermal densities of “prompt” ground-state charmed hadrons for hadronization temperatures of $T_H = 170$ MeV (including strong feeddowns) in the PDG and also added the one expected by the RQM. Numbers from [37].

verse momentum, indicating that at low momentum, the coalescence plays a significant role in the charm hadronisation procedure, where abundant lights quarks from the thermalised medium, whereas in high momentum region, fragmentation is the dominant contribution. Particularly, in this model, a condition $P_{coal}(p \rightarrow 0) = 1$ is a key aspect, meaning that charm quark can only hadroinse by coalescence at $p \simeq 0$. With this condition, it can predict a quite large values of heavy baryons. It is important to note that, the P_{coal} of Λ_c and Ξ_c is higher than $P_{coal}^{D^0}$, which explains the enhancement of the baryon-to-meson ratios reported in Fig. 1.9. In particular, this model foresees $P_{fragm} > P_{coal}$ for the charm mesons, and $P_{fragm} < P$ for the charm baryons given a hadron p_T lower than $10 \text{ GeV}/c$.

1.3.2.4 Statistical Hadronization Model with Relativistic Quark Model states

The statistical hadronization model (SHM) has been employed successfully in reproducing the yield of mesons and baryon ground states in Pb–Pb collisions at both RHIC and LHC energies, with one chemical freeze-out temperature of about $T_{cf} \approx 156 \text{ TeV}$ [115]. It assumes that hadrons are produced according to the density of thermal yield $N_h^{th}(T)$, and is governed by the masses of available hadron states at a universal hadronisation “temperature”, T_H . Such an approach has been able to capture the abundancies of D meson states, but is challenged by the recent large ratio of Λ_c^+/D^0 at the ALICE in pp collisions at $\sqrt{s} = 7 \text{ TeV}$ [116]. In fact, SHM only using charm-hadron state reported in PDG [45] can reproduce well the meson yields, but substantially underpredicted the charm baryons. However, it has been argued [37] that the charm baryon enhancement is due to the enhanced feed-down from “missing” charm-baryon states in the PDG.

The model Ref. [37] provides an updated framework that includes an augmented set of higher-mass charm baryons states predicted by the relativistic quark model (RQM) [117]. There are 18 extra Λ_c^+ states, 42 Σ_c ’s, 62 Ξ_c ’s, and 34 Ω_c ’s up to a mass of $3.5 \text{ GeV}/c^2$. As usual in the SHM, all the considered hadron states are governed by the thermal hadron densities with specific weights, following from their mass m_i and their spin degeneracy $d_i = 2J + 1$, evaluated at a hadronisation temperature T_H , as



$$n_i = \frac{d_i}{2\pi^2} m_i^2 T_H K_2\left(\frac{m_i}{T_H}\right), \quad (1.14)$$

where K_2 is the modified Bessel function of second order. The upper estimate of T_H is 170 MeV based on the results in Ref. [118].

The thermal densities are applied to normalise properly the charm-quark fragmentation functions, and to compute the charm p_T spectrum produced in pp collisions. All those additional excited charmed baryons mentioned above have to be forced to decay into a ground state. This procedure does not change the production of the ground state D mesons. However, including the additional baryon resonances makes a decisive difference and enables a significant charm baryon enhancement, as shown in Tab. 1.1.

1.3.3 Physics motivation

The measurement of open charm hadron production in pp collisions is a powerful tool for the test of pQCD calculations describing the production of charm quarks in hard scattering processes. In heavy-ion collisions, charm quarks are an excellent probe for the created QGP due to their large mass. Hence, the measurement of charm hadron production in these systems gives the opportunity to study and characterise the properties of the QGP.

As discussed earlier, the models introduced in Section 1.3.2 can provide a better description of Λ_c^+ (*cud*) production than $\Xi_c^{+,0}$ (*csu*, *csd*). This implies a possible difficulty in studying charm-strange baryons, demanding more differential measurements in the heavy-flavour baryon sector. Specifically, this thesis focuses on the study of charm-strange baryon production, to investigate the charm hadronisation mechanism.

A comprehensive explanation of Ξ_c^0 measured in pp collisions at $\sqrt{s} = 5.02$ TeV is detailed in Chapter 3. The measurement of Ξ_c^0 production in different collision energies is fundamental to understand the dynamics of heavy-flavour baryon formation, providing a key test for the different scenarios proposed in the mentioned models. This measurement also contributes to the computation of the total $c\bar{c}$ cross section in pp collisions at $\sqrt{s} = 5.02$ TeV, seen in Fig. 1.10. Moreover, it will serve as a reference and preparation for the planned measurements in Pb–Pb collisions to study the influence of a QGP, which is not yet feasible with the data currently available.

The Ω_c^0 baryon is composed of a charm quark and two strange quarks (*css*). The measurement of Ω_c^0 production plays a crucial role in constraining models and understanding whether strange quarks, or strange diquarks, have a special relevance for charm-baryon formation in pp collisions. Moreover, the

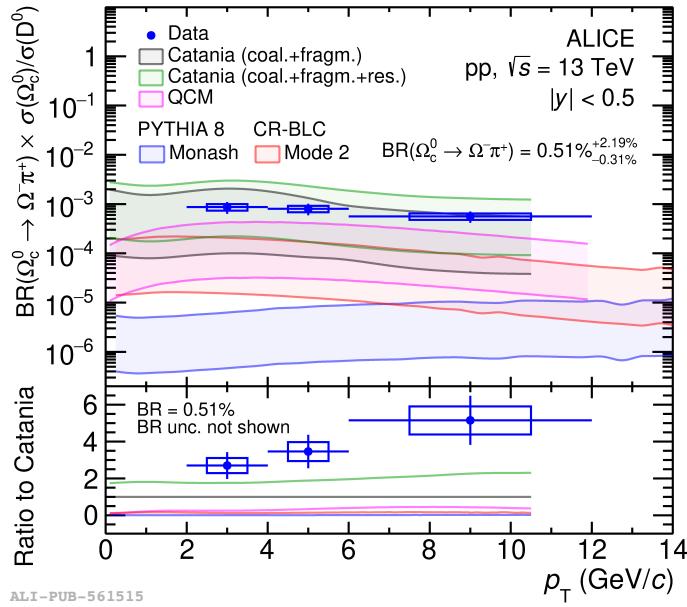


Figure 1.13: Top panel: ratio of the p_T -differential cross section of Ω_c^0 baryons (multiplied by the branching ratio into $\Omega^- \pi^+$) to the D^0 -meson one in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ [25]. Bottom panel: ratio of the data and models to the Catania (coalescence plus fragmentation) model [24].

comparison of the relative abundances of the $\Xi_c^{+,0}$ and Ω_c^0 in pp and e^+e^- collisions can shed light on the role of diquarks, because it can be sensitive to the different mass values of spin $S = 1(ss)$ and $S = 0(sd)$ diquarks [117].

Currently, a significant limitation in the interpretation of the production measurements of heavier strange-charm baryons is the absence of precise branching fraction values. In the recent inclusive $\Omega_c^0 \rightarrow \Omega^- \pi^+$ measurement in Ref. [25], shown in Fig. 1.13, model calculations were multiplied by a theoretically obtained branching ratio, $\text{BR}(\Omega_c^0 \rightarrow \Omega^- \pi^+) = (0.51^{+2.19}_{-0.31})\%$. This value was obtained by considering the estimate reported in Ref. [40] for the central value, and the envelope of the values (including their uncertainties) reported in Refs. [40, 119–123] to determine the uncertainty. This large uncertainty limits the understanding of Ω_c^0 production, therefore making it imperative to measure its branching ratios. Both the Belle and CLEO Collaborations published measurements of the branching-fraction ratio of the semileptonic and hadronic decay channels, $\text{BR}(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e) / \text{BR}(\Omega_c^0 \rightarrow \Omega^- \pi^+)$ and found 1.98 ± 0.13 (stat.) ± 0.08 (syst.) [39] and 2.4 ± 1.1 (stat.) ± 0.2 (syst.) [38], respectively. These measurements also test theoretical model calculations. Model calculations based on the light-front approach and on the light cone sum rules predict lower values of the branching fraction ratio of 1.1 ± 0.2 [40] and 0.71 [41], respectively. These models provide a way to relate the properties of hadrons, such as their masses, decay constants, and form factors, to fundamental QCD parameters and quark-gluon distributions within the hadrons. The hadronic part of the weak decay is parameterized in those models in terms of form factors,



which belong to the non-perturbative region of QCD. Another tool to study the charmed baryon decays is based on the flavor symmetry of $SU(3)_f$ in the quark model, which allows the calculations of decay modes and relative probabilities of charmed baryon decays [42, 124, 125]. By applying the effective colour approach under the framework of $SU(3)_f$ symmetry, a branching fraction ratio of 1.35 [42] is computed. The differences between the model calculations and the experimental values underscore the need for further experimental measurements and theoretical developments.

1.3.4 Organisation of the thesis

This thesis focuses on the study charm-baryon production in proton-proton collisions, as well as the charm quark hadronisation mechanism in the ALICE experiment at the LHC. The thesis is organised as follows:

Chapter 1 provides an introduction to the theoretical framework of the Standard Model, also a fundamental component of the Standard Model – the quantum chromodynamics (QCD) describing the strong interaction. Three distinctive features of QCD are listed in Section 1.2. Open heavy-flavour hadrons are most important tools to study the QCD theory in high-energy hadronic collisions, particularly in pp proton-proton collisions, where they offer sensitive tests for the calculations based on the perturbative QCD down to low p_T . The current status of open heavy-flavour in ALICE is discussed in detail in Section 1.3, with a primary emphasis on charm quark hadronisation, along with brief introduction to those developed models including those addressing baryon enhancement. The physics motivation of this thesis is also presented.

Chapter 2 begins with a brief introduction to the LHC experiment (Section 2.1), followed by a main focus on the ALICE experiment (Section 2.2). The most important detectors used for track, the Inner Track Systems (ITS) and the Time Projection Chamber (TPC), and for Particle Identification (PID), the TPC and the Time-of-Flight (TOF), which are used in these two analyses, are described in more detail in the following sections. It also outlines the track and vertex reconstruction procedures in ALICE in Section 2.3, as well as the ALICE offline analysis framework in Section 2.4. Additionally, an overview of the ALICE upgrade, ALICE 2 and ALICE 3, is given.

Chapter 3 presents the measurement of the p_T -differential cross section of Ξ_c^0 in pp collisions at $\sqrt{s} = 5.02$ TeV, including detailed analysis strategy: data sample, event selection, candidate selection and reconstruction, signal extraction, unfolding technique, reconstruction efficiency correction, and also corresponding systematic studies. The final results, compared with previously published measurements at $\sqrt{s} = 7$ TeV and 13 TeV, as well as with model predictions, are discussed in Section 3.6.



Chapter 4 focuses the p_T -differential cross section of Ω_c^0 baryons production multiplied by the branching ratio into $\Omega^- e^+ \nu_e$ measured in pp collisions at $\sqrt{s} = 13$ TeV, along with the branching-fraction $BR(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e)/BR((\Omega_c^0 \rightarrow \Omega^- \pi^+))$. To improve the precision of the measurements, there are three new techniques exploited for the Ω_c^0 analysis as explained in Section 4.1. A detailed explanation of raw yield extraction with the implemented BDT models, unfolding procedure, and efficiency correction is followed. The last part of this chapter discusses the results compared to those measured from CLEO and Belle Collaborations, as well as two theoretical calculations, seen in Section 4.8.

Chapter 5 summarises the two experimental measurements and presents some performance results from Run 3 in ALICE 2 and projection results from ALICE 3. In the end, the perspectives of new measurements that could be accessible in ALICE 2 and ALICE 3 are shortly discussed.



2 Experimental setup

In this chapter, a brief introduction to the Large Hadron Collider (LHC) will be presented. Following this, a detailed description of the ALICE (A Large Ion Collider Experiment) detector, with a particular focus on the main central barrel detectors relevant to this analysis, is shown. Then the vertex and track reconstruction, as well as the ALICE offline framework is shortly explained. Finally, the ALICE detector upgrade is briefly discussed.

2.1 The Larger Hadron Collider (LHC)

The Larger Hadron Collider (LHC) is the largest and most powerful high-energy particle accelerator, located at the European Organization for Nuclear Research (CERN), in Geneva. It lies in a tunnel 27 km ring of superconducting magnets and extends to depths of up to 175 meters. Primarily, the purpose of the LHC is designed for proton-proton ($p\bar{p}$) collisions, which allows us to investigate a wide range of fundamental questions in particle physics. Additionally, it can also accelerate heavy-ion beams, giving the chance to study extreme conditions like those present in the early universe, achieved through lead–lead ($Pb-Pb$) and $Xe-Xe$ collisions. It can also be used for proton-lead ($p-Pb$) collisions, to study the production mechanisms.

Fig. 2.1 shows a schematic illustration of the CERN’s accelerator complex with the four main LHC experiments. ATLAS (A Toroidal LHC Apparatus) [126] placed in site 1, and CMS (Compact Muon Solenoid) experiments [127] placed in site 5, are designed to measure the Higgs boson and explore the physics beyond the Standard Model. LHCb (Large Hadron Collider beauty) experiment [128] situated at site 8, is specialised in the study of heavy-flavour beauty quarks. ALICE (A Large Ion Collider Experiment) located in site 2, is dedicated to investigating heavy-ion collisions. Besides that, there is also the injector chain including for example LINACs and the SPS, designed to detect different phenomena. Detailed information about the LHC accelerator can be found in Ref.[129].

2.2 The ALICE experiment

The ALICE detector is a general-purpose heavy-ion detector [130], located at Interaction Point 2 of the LHC. It provides an unprecedented opportunity to investigate a small part of the phase diagram of strongly interacting matter and the physics of quark-gluon plasma (QGP) produced in heavy-ion collisions, both at extreme values of energy density and temperature. To address this, it requires the ability to

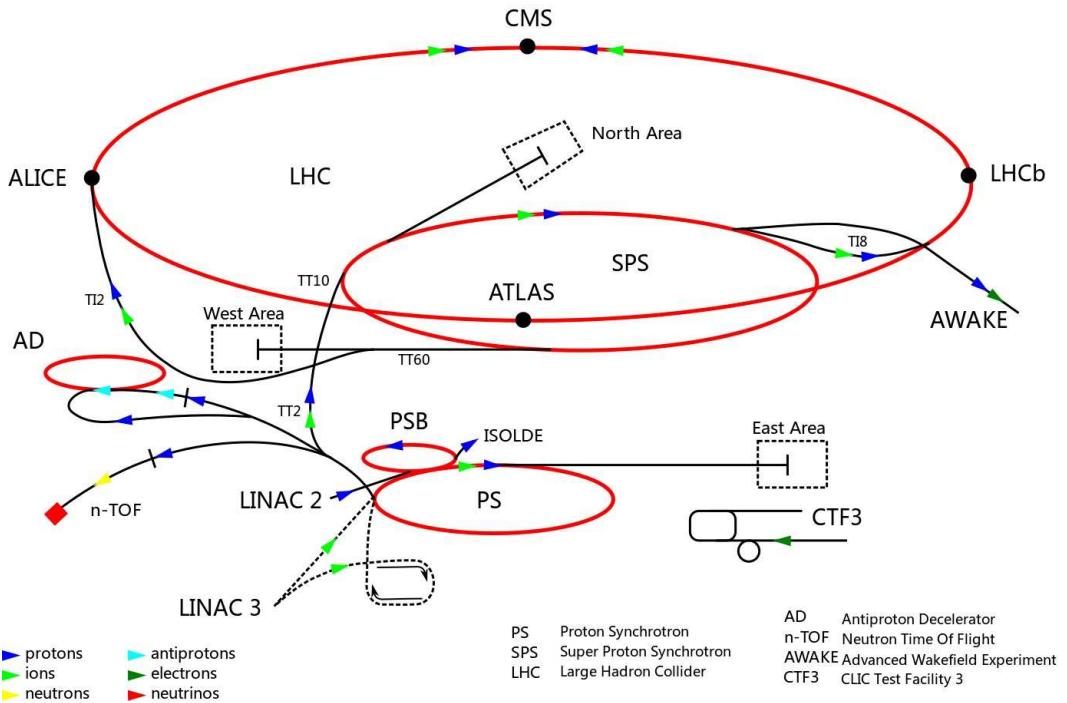


Figure 2.1: Schematic layout of the CERN accelerator complex including the LHC with the four big experiments and the injector chain [26].

measure charged particles in a wide momentum range, down to low transverse momenta with high precision (about $100 \text{ MeV}/c$ for pions with a relative momentum resolution of about 2% achieved [130]). The detector is also optimised to cope with high charged-particle multiplicity environments, ranging from 2000 to 8000 for central Pb–Pb collisions [130]. Furthermore, it provides excellent capabilities in tracking, for primary and secondary vertex reconstruction (i.e. $\sim 100 \mu\text{m}$ resolution for the study of the 2- and 3-prong decays of D mesons), and Particle Identification (PID).

The ALICE detector has overall dimensions of $16 \times 16 \times 26 \text{ m}^3$ and a substantial total weight of 10000 tons. An overview of the ALICE detector and its subdetectors is shown in Fig. 2.2. It can be divided into two main parts: the central barrel, covering the central rapidity region ($|\eta| < 0.9$), and the muon spectrometer positioned at forward rapidity ($-4 < \eta < -2.5$). The central barrel detectors are embedded in the L3 solenoid magnet, providing a maximum magnetic field strength of 0.5 T ($B = 0.5 \text{ T}$). In order of radii from the beamline, the central barrel detectors include the Inner Tracking System (ITS), the Time Projection Chamber (TPC), the Transition Radiation Detector (TRD), the Time of Flight (TOF), the Photon Spectrometer (PHOS), the Electromagnetic Calorimeter (EMCal) and High Momentum Particle Identification Detector (HMPID).

The ALICE coordinate system is a right-handed orthogonal Cartesian system. The x-axis is perpendicu-



THE ALICE DETECTOR

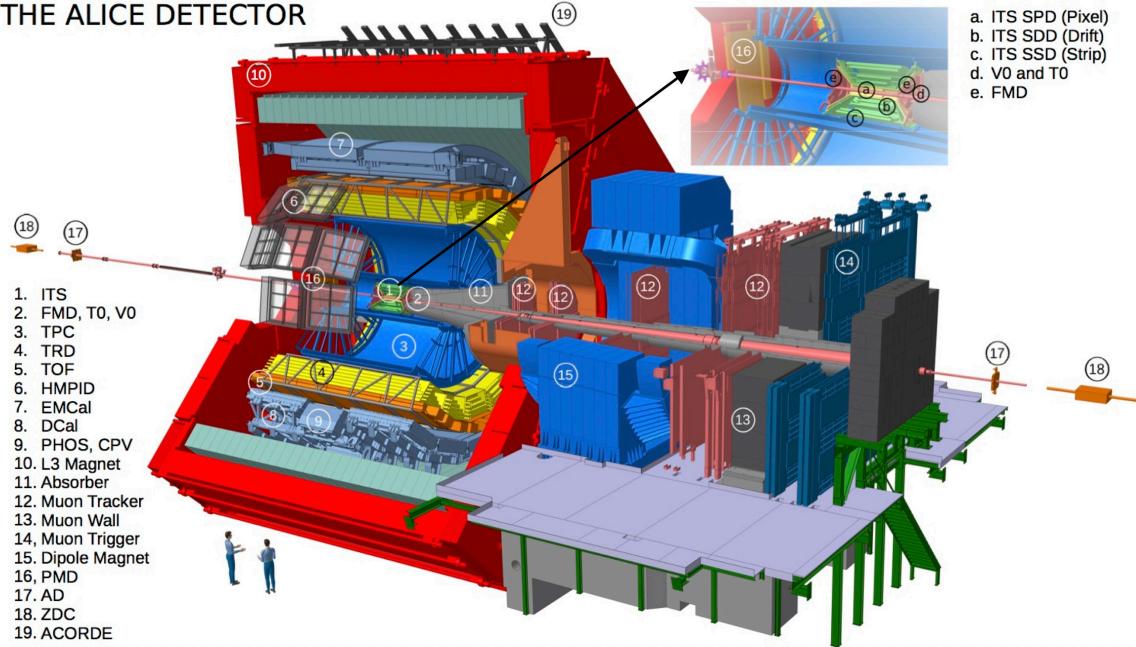


Figure 2.2: The ALICE experimental apparatus with the names of the sub-detectors. Figure from ALICE figure repository.

lar to the beam direction pointing to the centre of the accelerator, while the y-axis points upwards, and the z-axis is parallel to the beam direction along with the muon spectrometer in the negative z direction [130]. A detailed overview of the ALICE experiment and its performance can be found in Refs. [29, 130–132]. In the following sections, the detectors applied in this work are described.

2.2.1 Inner Tracking System (ITS)

The Inner Tracking System (ITS) is an important part of the tracking system, positioned closest to the beam pipe and the interaction point. It comprises six cylindrical layers of silicon detectors, depicted in Fig. 2.3. It covers the pseudorapidity region $|\eta| < 0.9$. These six layers employ various detector technologies to achieve their functions. Detailed information about each of the layers is provided in Tab. 2.1.

Table 2.1: Brief summary of the different layers of the ALICE ITS.

	SPD	SDD	SSD
Distance from beam pipe	3.9 and 7.6 cm	15.0 and 23.9 cm	38.0 and 43.0cm
$r\varphi$ precision	$12 \mu m$	$35 \mu m$	$20 \mu m$
z precision	$100 \mu m$	$25 \mu m$	$830 \mu m$
Readout	Digital	Analog	Analog

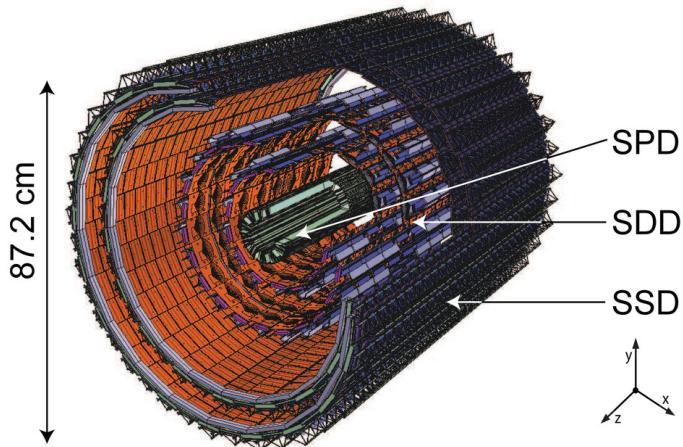


Figure 2.3: Structure of the ITS, Figure from [27].

The primary responsibilities of the ITS include:

- Reconstructing the primary vertex of collisions with a high resolution, surpassing $100 \mu\text{m}$
- Reconstructing the secondary vertices, points of decay for short-lived hadrons
- Particle identification and tracking of low-momentum particles with momentum below $200 \text{ MeV}/c$

The innermost 2 layers of the ITS detector are constructed using hybrid pixels known as Silicon Pixel Detectors (SPD). One of the main functions is to determine the position of the primary vertex, by selecting pairs of reconstructed points in these two layers. The resolution of the primary vertex position depends on the track multiplicity, in turn, which is related to the charged-particle density. The extended pseudo-rapidity coverage of the SPD ($|\eta| < 1.98$) ensures the accurate determination of track multiplicity. Furthermore, their fast response makes them an integral part of the overall ALICE triggering system, facilitating rapid event selection.

The two intermediate layers are the so-called Silicon Drift Detector (SDD). The SDD detector has very good multi-track capability and provides two out of the four measurements of specific dE/dx employed for PID by the ITS. However, in this thesis, the latter feature is not used, due to other detectors with excellent PID capabilities, which will be discussed later in Section 2.2.3.

The outermost two layers employ double-sided Silicon Strip Detectors (SSD). The SSD detector is essential for the prolongation of tracks from the main tracking detector of ALICE, the Time Projection Chamber explained in Section 2.2.2, to the ITS. Besides that, it also provides dE/dx information, which enables the ITS with the standalone capability to serve as a low- p_T particle spectrometer.



2.2.2 Time Projection Chamber

The Time Projection Chamber (TPC) [28], illustrated in Fig. 2.4, is the main tracking detector of the central barrel. The TPC is specifically designed to measure tracks with transverse momentum (p_T), ranging from $0.1 \text{ GeV}/c$ to $100 \text{ GeV}/c$, and also has excellent PID capabilities via the information of dE/dx . The TPC has the shape of a hollow cylinder with a radius ranging from 85 cm to 250 cm and an overall length along the beam direction of 5 m, covering the full azimuthal angle. It covers the pseudorapidity range of $|\eta| < 0.9$ for tracks with full radial track length (matches in ITS, TRD and TOF), and up to $|\eta| < 1.5$ for slower particles (reduced track length).

The detector is divided along the beam axis into two equally large drift regions by a central cathode, each having a length of 2.5 m. On the opposite sides of the central electrodes is the readout, which is equipped with multi-wire proportional chambers (MWPC). Each of the 18 readout chambers covers an azimuth of 20° . The interior of the TPC is filled with a NeCo_2N_2 or ArCO_2N_2 gas mixture with a small radiation length and low multiple scattering rates. During Run 2 the gas was changed from a Neon mixture to an Argon mixture and back again due to the large observed space-charge distortions.

Charged particles passing through the gas volume ionize the gas, resulting in the liberation of electrons. These free electrons drift toward the end plates and hit the readout pads, under the influence of a highly uniform electrostatic field of $400 \text{ V}/\text{cm}$ in the beam direction. The signal amplification of the freed electrons is provided through the avalanche effect occurring in MWPCs. Due to the influence of the magnetic field, the particles follow helical paths as they traverse the TPC. The momentum of charged

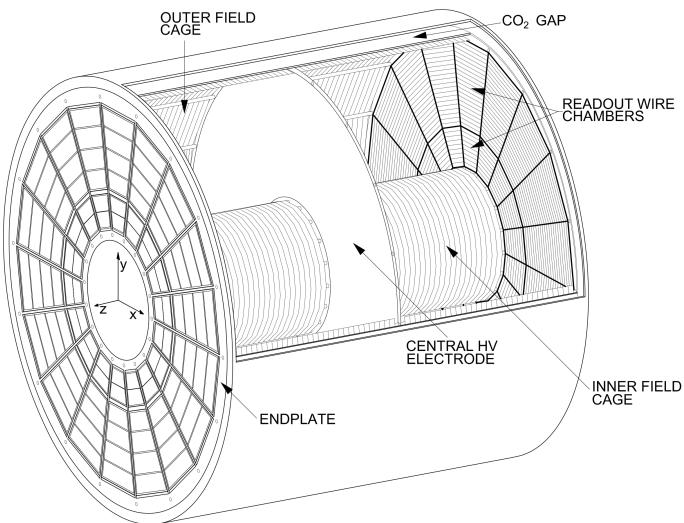


Figure 2.4: Layout of the Time Projection Chamber. Figure from [28].

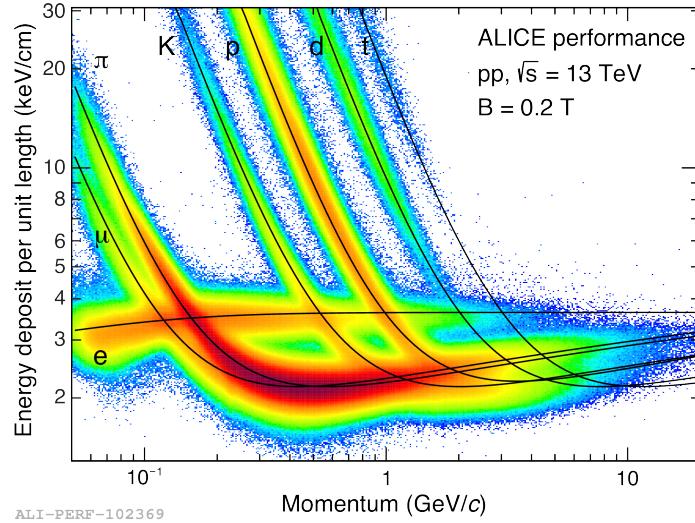


Figure 2.5: Charged particles dE/dx as a function of momentum measured by the TPC in pp collisions at $\sqrt{s} = 13$ TeV. Figure from ALICE Figure Repository.

particles can be measured based on the radius of curvature of its track in the magnetic field. The track position in $r\phi$ can be determined by the pad position of the deposited charge, where the z -direction is calculated from the drift time of the electrons. This is how the information for a full 3D reconstruction of the tracks is obtained.

The TPC is equipped to provide important PID information based on the charge and momentum of particles, as well as the measurement of the specific energy loss (dE/dx) associated with the process of ionization. The specific energy loss can be described by the Bethe-Bloch formula, which depends on the particle species and its momentum, and the properties of the traversed medium. The PID determination is estimated by comparing the measured energy loss in the TPC with the expected dE/dx of a specific particle species and its momentum. The expected energy loss is parameterised by a function originally proposed by the ALEPH collaboration [133], which is a fundamental tool for particle identification in the experiment,

$$f(\beta\gamma) = \frac{P_1}{\beta^{P_4}} \times \left(P_2 - \beta^{P_4} - \ln(P_3 + \frac{1}{(\beta\gamma)^{P_5}}) \right), \quad (2.1)$$

where β is the particle velocity, γ is the Lorentz factor, and P_{1-5} are parameters from fits to measured data. A specific particle species i can be identified by the deviation of the measured dE/dx from the expected energy loss in terms of the measurement resolution $\sigma_{dE/dx}$,



$$n_{\sigma}^i = \frac{\frac{dE}{dx}_{\text{TPC}} - \langle \frac{dE}{dx} i \rangle}{\sigma_{dE/dx}}, \quad (2.2)$$

where $\langle dE/dx \rangle$ is directly proportional to the charge of the electrons released during ionisation and is collected from the readout pads within the TPC. This allows TPC to identify various particle species. The measured dE/dx in the TPC as a function of momentum in pp collisions at $\sqrt{s} = 13$ TeV is shown in Fig. 2.5. The solid lines are the expected trends for different particle species, according to the Eq. 2.1.

The TPC has excellent PID ability via dE/dx determination with a resolution better than 5% [28]. At low momenta ($p \leq 1$ GeV/c) particles can be identified on a track-by-track basis, while at higher momenta, particles can be still separated on statistical methods via multi-Gaussian fits. To cover tracks up to a few GeV/c, the PID information has to be complemented by additional measurements provided by other subdetector systems.

2.2.3 Time Of Flight

The Time Of Flight (TOF) detector is a large array of Multi-gap Resistive-Plate Chambers (MRPC), located at 3.7 m and 3.99 m radial distance, respectively, from the interaction point. The TOF detector design is schematically shown in Fig. 2.6. It is designed to identify charged particles produced in the pseudo-rapidity of $|\eta| < 0.9$ and within the intermediate momentum range ($0.5 < p_T < 4$ GeV/c). If a particle ionises the gas in the detector, an avalanche process will be triggered to generate signal on the readout electrodes. The total signal is the analogue sum of signals from many gaps and the time resolution is about 40 ps with a detector efficiency of close to 100% [29].

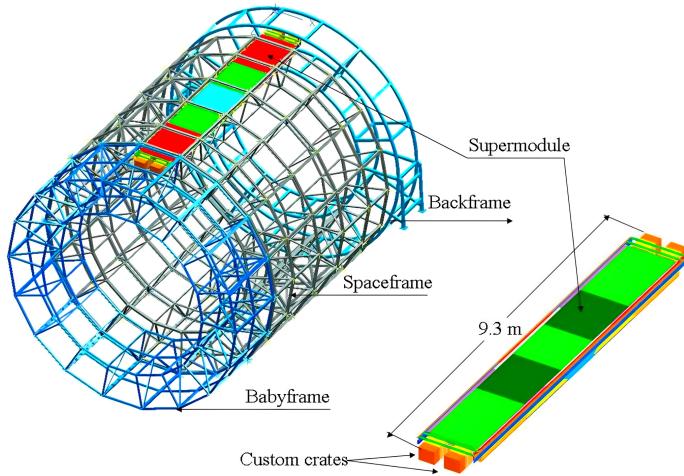


Figure 2.6: Modular structure of the ALICE TOF detector. Figure from [29]

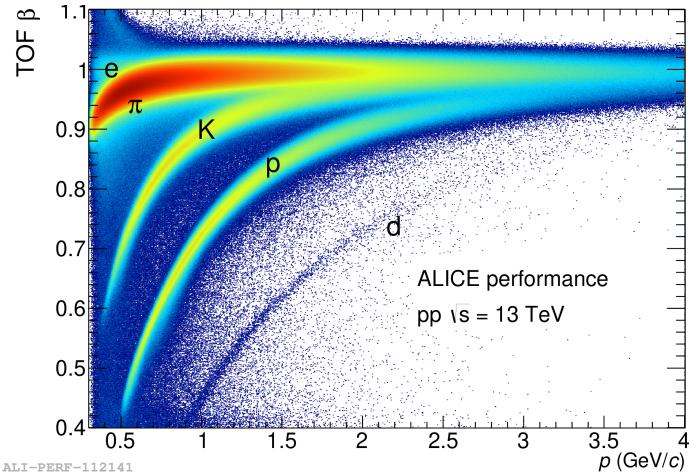


Figure 2.7: β of charged particles as a function of momentum measured by the TOF in pp collisions at $\sqrt{s} = 13$ TeV. Figure from ALICE Figure Repository.

The particle is identified by the measurement of particle time-of-flight relative to the time of collisions, over a given distance along its trajectory L and its momentum, which is known from the TPC. The velocity of the particle is given by, $\beta = v/c = L/tc$. One can determine the mass and thus identify the particle by,

$$m = \frac{p}{c} \sqrt{\frac{c^2 t^2}{L^2} - 1} \quad (2.3)$$

The identification of particles using the TOF detector relies on the number of standard deviations (n_σ), in the time of flight difference between two particles with masses m_A and m_B with same momentum p :

$$n_\sigma = \frac{|t_A - t_B|}{\sigma_{\text{TOF}}} = \frac{Lc}{2p^2\sigma_{\text{TOF}}} |m_A - m_B|, \quad (2.4)$$

where $\sigma_{\text{TOF}} \approx 60 \sim 80$ ps represents the overall time resolution of the TOF detector [29]. This relation is valid if $p \gg mc$.

The PID capabilities measured by the TOF detector in pp collisions at $\sqrt{s} = 13$ TeV is shown in Fig. 2.7. TOF can effectively separate pions and kaons below $2.5 \text{ GeV}/c$ and provide the separation for protons and kaons up to $4 \text{ GeV}/c$. Furthermore, the information from the TOF also can be utilized to identify electrons and nuclei. The TOF helps the PID in the intermediate momentum region, along with the ITS and TPC detectors.



2.3 Track and vertex reconstruction

In ALICE, the track-reconstruction and the vertex-finding procedures are performed offline. The tracking procedure starts with the *clusterization step*. First, for each detector, the collected raw data are transformed into *clusters*, which correspond to groups of hits produced by a single particle interaction with a detector element. The clusters are characterized by their positions, signal amplitude, signal times, and their associated errors. The clusterization is performed separately for each detector.

The reconstruction of the primary vertex is an essential step, as it is referred to the interaction point. Usually, each collision event only involves one primary interaction point, and numerous tracks traverse the detectors. The information provided by the SPD detector plays a crucial role in establishing a preliminary interaction vertex. This is the space point that minimises the distance among the *tracklets*, which are the track segments reconstructed by associating pairs of clusters in the two SPD layers. Then the vertex is defined as the point where most of the tracklets converge. In high multiplicity events, where interaction pileup is expected, the algorithm is repeated several times, discarding at each iteration those clusters which contributed to already-found vertices. The pile-up events (less than 1%) are rejected by detecting multiple primary vertices using track segments defined with the SPD layers. In case no single convergence point is found, particularly in low-multiplicity pp collisions, the algorithm performs a one-dimensional minimisation in the z -distribution of the points of closest approach (PCA) of tracklets to the measured beam-axis position. This measurement of the primary-vertex position is used as an input for the tracking.

The track reconstruction procedures normally involve fit problems, which can be solved by using the Kalman filter (KF) vertexing algorithm [134]. Especially, it is implemented in a three-pass tracking scheme known as *inward-outward-inward*, as shown in Fig. 2.8. This method depends critically on the determination, for each track, of a set of initial seed values for track parameters and their covariance matrix. A brief introduction about the KF is presented in Section 4.1.1.

The first inward stage starts with finding tracks in the TPC at a large radius. As a first step, the track seeds are built from the clusters of the two outermost pad rows of the TPC and the primary vertex point as a constraint, and then using three clusters without any constraint to the vertex. This approach allows for the efficiency track seeds propagation until reaching to the TPC inner radius, meanwhile, they are updated at each step by assigning clusters that fulfill certain proximity cuts by the KF algorithm. Clusters can be reused by multiple seeds, it is possible to reconstruct the same physical track multiple times. To mitigate this issue, a dedicated algorithm is used to identify pairs of tracks that share a fraction of common clusters above a certain threshold, typically ranging from 25% to 50%. Additionally, only those

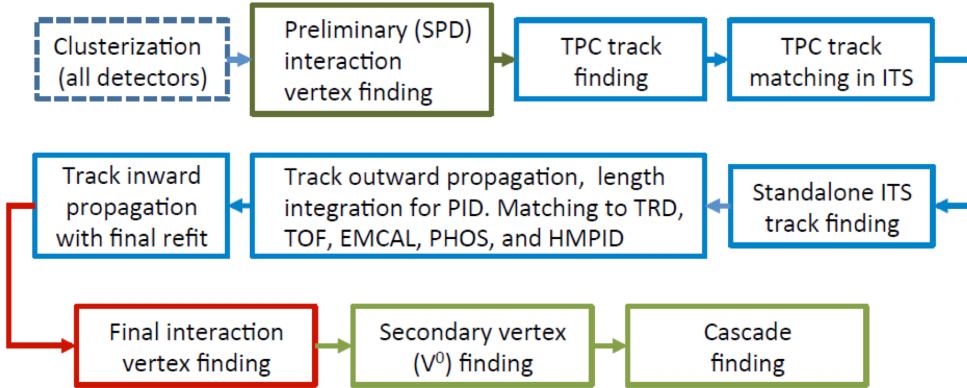


Figure 2.8: Event reconstruction flow.

tracks with at least 20 out of 159 clusters in the TPC are accepted for reconstruction. These are then propagated inwards to the inner TPC radius.

The efficiency of reconstructed tracks in the TPC as a function of p_T , is shown in the left panel of Fig. 2.9, in pp collisions at $\sqrt{s} = 8$ and Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. At low transverse momentum (~ 0.5 GeV/c), it drops dramatically caused by the energy loss and multiple scattering within the detector material. The shape of the efficiency at high p_T is due to the loss of clusters in the dead zones between the TPC sectors.

The reconstructed tracks from the track-finding step in the TPC are then propagated to the outer layer of the SSD, which becomes the seeds for the ITS track finding. A similar procedure is adapted, tracks are propagated inward and updated at each ITS layer attaching all clusters within a proximity cut. Each TPC

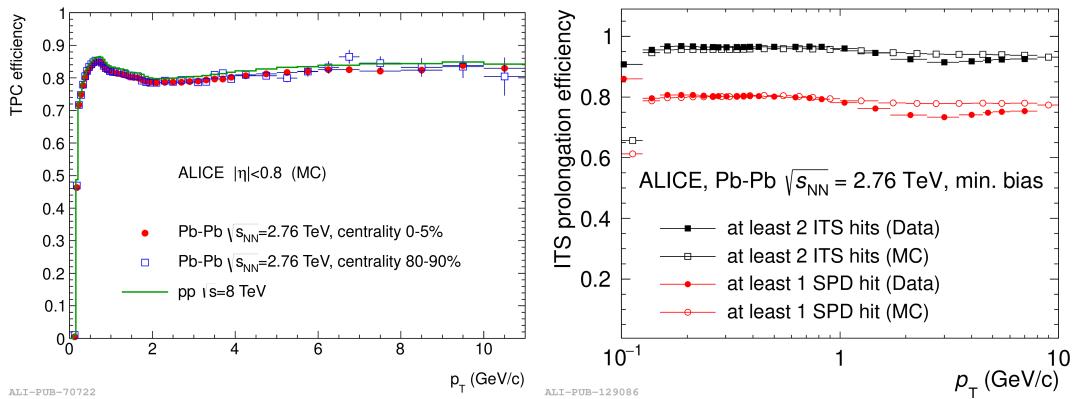


Figure 2.9: Left panel: reconstruction efficiency for TPC tracks in pp collisions at $\sqrt{s} = 8$ TeV and in central and peripheral Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Right panel: matching efficiency of tracks from ITS to TPC in real (full markers) and simulated (open markers) Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Figures from [29].



track produces a tree of track hypotheses in the ITS and candidates are stored based on the reduced χ^2 . An algorithm is applied to avoid two tracks sharing common clusters, in order to determine the tracks with the highest quality, and then they are added to the reconstructed event. The TPC-ITS matching efficiency in this step is defined as the ratio of the fraction of TPC tracks prolonged in the ITS. The right panel of Fig. 2.9 shows the matching efficiency in data and the simulation for Pb–Pb collisions at $\sqrt{s_{\text{NN}}} = 2.76$ TeV, with different requirements on ITS points [29].

To increase the efficiency for tracks with low $p_T < 200$ MeV/c and to reconstruct tracks not hitting in the TPC (traversing dead zones of the TPC or decay before entering the TPC), an ITS stand-alone reconstruction procedure is performed by using seeds created from two clusters originating in the three innermost ITS layers and the primary vertex point.

After ITS tracking is completed, all tracks are extrapolated to their point of closest approach to the preliminary interaction vertex. The tracking-procedure second step starts, the obtained tracks are then propagated outwards using the KF algorithm. Once the tracks reach the TRD, an attempt is made to match it with a TRD tracklet in each of the six TRD layers. Similarly, tracks arriving at the TOF detector are matched to the TOF clusters. Subsequently, the tracks are propagated further to match signals in EMcal, PHOS and HMPID with a requirement for minimum-ionising clusters.

In the last stage, the reconstructed tracks are propagated starting from the outer TPC radius toward the interaction point. In each detector (TPC and ITS), the tracks are refitted with the previously found clusters. After the full tracking procedure, the final track parameters and covariance matrix are computed. In case two or more tracks are reconstructed with the procedure described above, the final position of the primary vertex is recalculated using ITS-TPC tracks, which are also called *global tracks*. This method

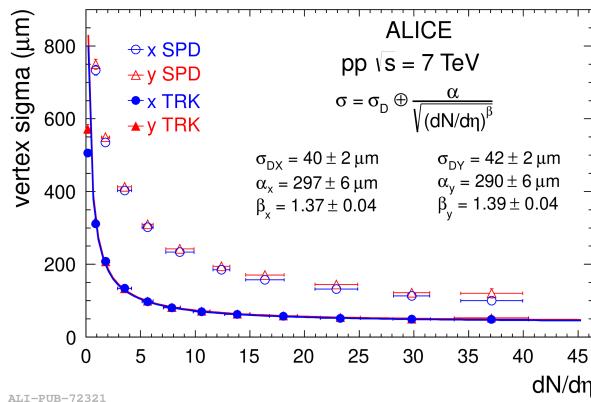


Figure 2.10: Resolution on the primary-vertex position using the SPD and the track algorithms as a function of the charged-particle multiplicity in pp collisions at $\sqrt{s} = 7$ TeV. Figure from [29].



improves the resolution by about a factor of 2.5 with respect to the preliminary primary vertex. From Fig. 2.10, the improvement of the resolution on the primary-vertex position obtained with global tracks concerning that obtained only with SPD tracklets is illustrated.

2.4 The ALICE offline framework

The ROOT system [135], being developed at CERN since 1995, is a cross-platform framework designed specifically for storage and data analysis in an efficient way. Nowadays, ROOT is widely used tool in the high-energy physics, and also gained users from other fields of science. ROOT is the core of the ALICE offline framework, having the two main parts:

- *AliROOT*: it is used for Monte Carlo simulations, data reconstruction, calibrations and detector alignment. It provides an interface to several event generators, such as PYTHIA [13] for pp collisions, used for Monte Carlo simulations. It includes a comprehensive description of the detector geometry and material budget information, implemented in independent modules aiming for different subsystems. To simulate the detector response, AliROOT transports to other packages such as GEANT3 [136], GEANT4 [137] and FLUKA [138].
- *AliPhysics*: It contains all the code used in the physics analysis. This is based on multiple tasks with a predefined structure, which are designed to process the collected and simulated events stored in the grid.

The ALICE offline framework will be completely renewed in view of the data taking campaigns in Run 3 and 4 at the LHC, being fully replaced by the *mO²* framework [139], which will be not described in this thesis.

After reconstruction, the events are called Event Summary Data (ESD) which contain all the physical information required for analyses and quality assurance for each subdetector. But, ESD files are of huge volume and not efficient to handle. Therefore, the data files are compressed to the Analysis Object Data (AOD) which are derived from ESD by filtering. Tracks that pass some sets of cuts are stored and the rest are deleted from AOD. Thus AOD contains only the physics-related data and is efficient to handle, though analysis can be done using both AOD and ESD.



2.5 ALICE detector upgrade

The ALICE experiment conducted during Runs 1 and 2 at the LHC, from 2015 to 2018, provide valuable results, leading to a deep understanding of the properties of the QGP. Despite the significant progresses that have been made so far, there are still several fundamental questions to be studied, and it will be not possible to fully answer these questions with the present detectors. Therefore, it is very necessary to upgrade the detector, to extend the physics cases and performances of the ALICE experiment. In the following, a very brief introduction about ALICE upgrade 2 and 3 is introduced, as well as the improvements.

2.5.1 ALICE 2 - Run 3

During the LHC long shutdown 2 (LS2) from 2019 to 2021, a major upgrade of the experimental apparatus was installed in the ALICE detector, preparing for the upcoming larger interaction rates expected during Run 3 (2022 - 2025) and 4 (2029 - 2032) of Pb–Pb collisions. This detector upgrade referred to as *ALICE 2*, following specifications outlined in Ref. [140]. The key changes implemented include: 1) the upgrade of the TPC readout chambers replaced with Gas Electron Multiplier (GEM) chambers [141], enabling continuous readout of Pb–Pb events at an interaction rate up to 50 kHz while maintaining momentum measurement and particle identification performance as mentioned above; 2) new vertex detector ITS 2 [30], comprising seven layers equipped with Monolithic Active Pixel Sensors (MAPS), with a smaller-diameter beam pipe to enhance track and vertex resolution; 3) new apparatus of Muon Forward Tracker (MFT) [142], featuring five double-side detection disks for precise tracking and secondary vertex reconstruction for muon tracks; 4) a new Fast Interaction Trigger detector (FIT), based on Cherenkov radiators and scintillator for fast event triggering [143]. The upgraded detector is designed for high readout rates of 50 kHz in Pb–Pb (with respect to the previous few kHz) and 1 MHz in pp collisions. These improvements enable more precise measurements, especially at low p_T . Moreover, an entire new Online-Offline (O^2) software framework [139] was deployed, together with the implementation of the First Level Processors (FLP) and the Event Processor Node (EPN) computing nodes, in order to cope with the much higher data throughout the data taking.

A significant aspect of the upgrade is the installation of the ITS 2 in the central barrel, with the Inner Barrel (Layers 0 to 2) and the outer Barrel (Layers 3 to 6), as shown in Fig. 2.11. The smaller inner radius of the newly installed beam pipe (18.2 mm) brings the inner layers closer to the beam axis, leading to improvement of the resolution of the track impact parameter. In addition, the reduction of material budget in the first detection layer, about $0.36\%X_0$, significantly enhances the impact parameter resolution, comparing the one in ITS1 about $1.14\%X_0$ [31]. The improvement of the resolution in the distance of

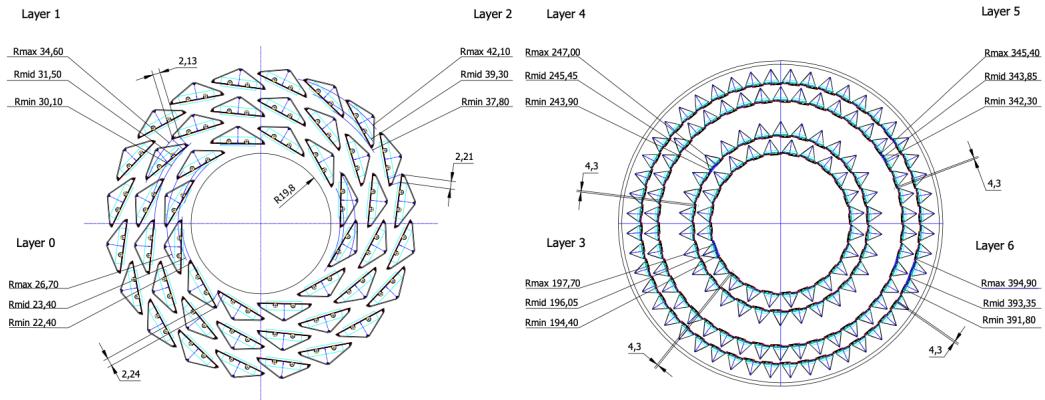


Figure 2.11: Schematic view of the cross section of the Inner Barrel (left) and Outer Barrel (right). Figure from [30].

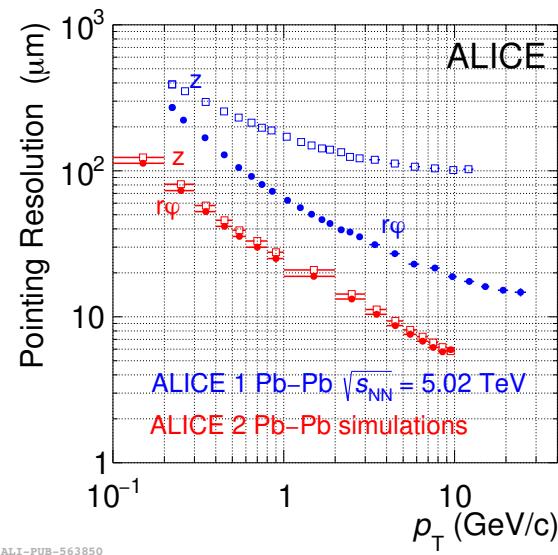


Figure 2.12: Transverse (solid circle markers) and longitudinal (open square markers) impact parameter resolution in Pb–Pb collisions data from ALICE 1 in blue points, and simulations from the upgraded detector ALICE 2 in red points. Figure from [31].

closest approach of tracks to the primary vertex in both the transverse ($r\varphi$) and longitudinal (z) directions in Pb–Pb collisions, obtained from the simulation with the full O² simulation framework for the upgraded detector in Run 3 with respect to the ones from Run 2 data, is illustrated in Fig. 2.12.

The second major upgrade involves the readout system of the TPC. The new TPC chambers based on gas electron multiplication foils, enable continuous readout at much higher interaction rates and reduce ion backflow to prevent large charge accumulation and space-charge distortions. The performance of the upgraded TPC demands complex and innovative calibration, and also an efficient data compression system. The upgraded TPC will offer a similar momentum resolution as the Run 2 system.

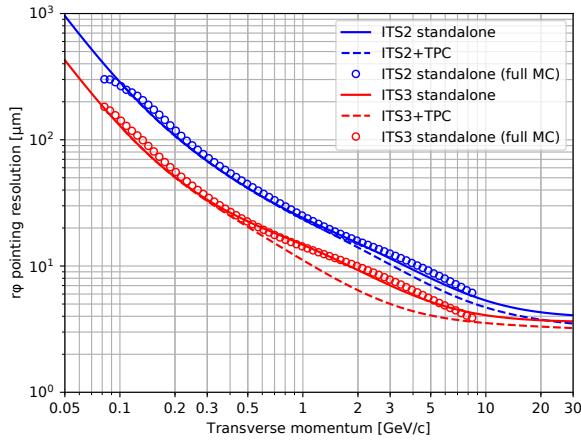


Figure 2.13: Impact parameter resolution for primary charged pions as a function of the transverse momentum for the ITS3 upgrade and ITS2 upgrade in the transverse plane. The solid lines represent the result with Fast Monte Carlo Tool (FMCT) and ITS only, the dashed lines are the results from FMCT, together with ITS+TPC, and open circles show the results with full MC and ITS only. Figure from [32].

ALICE 2 aims to improve its event-rate capabilities and enhance significantly the tracking and vertexing performance, particularly at low momentum. It has been designed with a focus on precise heavy-flavour measurements as one of the main physics goals. The expected projections from Run 3 regarding charm baryon measurements can be found in Chapter 5.

2.5.2 ALICE 2 - Run 4

Two additional upgrades are preparing for the Long Shutdown 3 [32, 144], from 2026 - 2028, to further enhance the physics of the experiment in Run 4. These upgrades include a new inner barrel for the Inner Tracking Systems, *ITS 3 project*, and the *FoCal project*, which employs the installation of a forward calorimeter optimised for photon detection in the rapidity range $3.4 < \eta < 5.8$. The ITS3 consists of three cylindrical layers utilising curved wafer-scale stitched sensors. These layers will be positioned even closer to the interaction point, with the first layer located just 18 mm away. Additionally, a new beam pipe with a smaller radius (inner radius of 16 mm) and reduced thickness ($500 \mu\text{m}$) will be installed, compared to the upgrade during LS2 (with an inner radius of 18.2 mm and a thickness of $800 \mu\text{m}$). This upgrade will result in an unprecedented reduction of the material budget by $0.05\% X_0$ near the interaction point. A comparison of the impact parameter resolution for primary charged pions at typical pseudorapidity $\eta = 0.5$ between ITS2 and ITS3 is shown in Fig. 2.13, where it is improved by a factor of two for $p_T = 1 \text{ GeV}/c$. Furthermore, the tracking efficiency at very low transverse momentum is also improved [32]. These improvements will have a substantial impact on the measurement of charm and beauty hadrons at low transverse momentum, as well as on the measurement of low-mass and low p_T dielectrons.



The Focal detector [144] consists of a high-granularity electromagnetic calorimeter combined with a conventional sampling hadronic calorimeter, located outside the ALICE solenoid magnet at a distance of 7 m from the interaction point. The electromagnetic part of Focal is a compact silicon-tungsten (Si + W) sampling calorimeter with longitudinal segmentation. Its main physics goal is to study the non-linear QCD evolution at small Bjorken- x down to $x \sim 10^{-6}$. Notably, benefiting from its excellent reconstruction capabilities, FoCal achieves a significantly better signal-to-background ratio compared to LHCb's photon conversion-based approach, as demonstrated in Fig. 43 of Ref. [144].

2.5.3 ALICE 3 - Run 5 and 6

The Long Shutdown 4 is foreseen from 2033 - 2034. The *ALICE 3*, a completely new setup for the LHC Runs 5 and 6 has been proposed, which enables new measurements in the heavy-flavour sector, focusing on low- p_T , especially for the measurements of multi-charm baryon production, multi-differential measurements of dielectron emission, as well as femtoscopic studies of the interaction potentials between heavy mesons.

The proposed novel ALICE 3 detector with a high readout rate, consists of a feature state-of-the-art tracking system with a unique pointing resolution across a broad pseudorapidity range ($-4 < \eta < +4$), complemented by multiple sub-detector systems for particle identification, reaching a momentum resolution of 1 - 2%. They are silicon TOF layers, a ring-imaging Cherenkov detector with high-resolution readout, a muon identification system, as well as an electromagnetic calorimeter. To optimise the pointing resolution, the first tracking layer must be placed as close as possible to the interaction point. One of the key innovations of ALICE 3 is its retractable vertex detector, designed to provide a larger aperture required for beams at the injection energy. This feature enables the detection of heavy flavour hadrons with exceptional purity and efficiency down to low p_T . Additionally, advancements in pointing resolution, projected to be approximately three times better than that of ITS3, offer unprecedented precision in heavy-flavour measurements at midrapidity.

The evaluated performance study of the multi-charm baryons in Run 5 with ALICE 3 is reported in Chapter 5.



3 Ξ_c^0 production in pp collisions at $\sqrt{s} = 5.02$ TeV

Measurements of heavy-flavour production in proton-proton (pp) collisions at LHC energies are among the most important tool to test the quantum chromodynamics calculations with the factorisation approach, as discussed extensively in Section 1.3.

In this chapter, the p_T -differential production cross section of prompt Ξ_c^0 via the semileptonic decay channel $\Xi_c^0 \rightarrow e^+ \Xi^- \nu_e$ and its charge conjugate, in pp collisions at $\sqrt{s} = 5.02$ TeV is presented. It is organised as follows. Section 3.1 describes the data samples and the data-taking conditions employed in this analysis. The Ξ_c^0 candidates are built by pairing an electron-candidate with Ξ candidate, as seen in Section 3.2. The unfolding procedure is implemented to correct for the missing momentum of neutrino, as explained in Section 3.3.3. Section 3.4 describes the acceptance-times-efficiency and prompt fractions, respectively. In Section 3.5, the evaluation of the systematic uncertainties for this analysis is discussed. Finally, the results will be presented and discussed, including the production of prompt Ξ_c^0 and charmed baryon-to-meson yield ratio Ξ_c^0/D^0 reported in Section 3.6, as well as the comparisons with model calculations.

3.1 Data sample and offline event selection

This analysis uses a Minimum Bias (MB) triggered data sample collected from pp collisions at the LHC in 2017 with a centre-of-mass energy of $\sqrt{s} = 5.02$ TeV. Events are selected with the MB trigger, which is the trigger configuration imposing the smallest bias by accepting most of the events. Not all trigger signals are caused by genuine collisions, there are some fake triggers generated by the background from the LHC beam. In pp collisions, offline event selections are applied to remove the contamination from interactions between the beams and the remnant gas in the beam pipe of the LHC. This selection is based on the timing information of the two V0 arrays [145] and a cut on the correlation between clusters and tracklets reconstructed in SPD. In order to maintain a uniform ITS acceptance in pseudorapidity, only events with a reconstructed vertex position within ± 10 cm along the beam axis from the nominal interaction point are analysed. After the aforementioned selections, the data sample used for the analysis consists of about 990 million MB events, corresponding to an integrated luminosity of $\mathcal{L}_{\text{int}} = (19.3 \pm 0.4) \text{ nb}^{-1}$ [146].

The Monte Carlo (MC) sample is generated with PYTHIA 8 event generator [13] and the GEANT 3 transport code [136], where $c\bar{c}$ pairs are generated in each event, while the presence of Ξ_c^0 is used as the condition to store the event. The simulations are anchored to the data-taking runs, as to reproduce the



detector and data-taking environment. There are about 50 million MC events used in this analysis.

3.2 Raw yield extraction

3.2.1 Electron identification

The electron candidates used to reconstruct Ξ_c^0 are required to meet several selections, which are summarized in Tab. 3.1. Only tracks with $|\eta| < 0.8$, $p_T > 0.5\text{GeV}/c$, at least 100 out of 159 associated crossed TPC pad rows, a χ^2 normalised to the number of associated TPC clusters smaller than 4, a ratio of crossed rows to findable clusters in the TPC greater than 0.6, and a minimum number of three hits (out of six) in the ITS, and both hits in the SPD, are considered.

These selection criteria suppress the contribution from short tracks, which are unlikely to originate from the primary vertex. Electrons are identified using the dE/dx and the time-of-flight measurements in the TPC and TOF detectors, described in Section 2.2.2 and 2.2.3, respectively. The selection is applied on the $n_{\sigma,e}^{\text{TOF}}$ and $n_{\sigma,e}^{\text{TPC}}$ variables defined as the difference between the measured dE/dx or time-of-flight values and the ones expected for electrons, divided by the corresponding detector resolution. The TOF detector in ALICE is capable of reducing the amount of hadron contaminations. Combined with the PID capability of the TOF detector, the TPC detector can reach a high electron purity.

In the top left panel of Fig. 3.1, the $|n\sigma_{\text{TPC}}^{\text{ele}}|$ distribution as a function of the candidate electron p_T is

Table 3.1: Electron identification cuts applied in this analysis.

Cuts variables	cuts
AOD Filter Bit	4 (Standard cuts with very loose DCA)
Number of TPC clusters	> 100
Number of TPC PID clusters	> 80
Ratio of findable clusters	> 0.6
ITS/TPC refit	TRUE
Number of ITS cluster	≥ 3
$p_T(\text{GeV}/c)$	> 0.5
$ \eta $	< 0.8
SPD hit	both
$ n\sigma_{\text{TOF}}^{\text{ele}} $	< 3
$ n\sigma_{\text{TPC}}^{\text{ele}} $	$> -3.9 + 1.17 \times p_T^{\text{ele}} - 0.094 \times p_T^{\text{ele}} \times p_T^{\text{ele}}$
prefilter cut	$m_{e^+e^-} < 0.05 \text{ GeV}/c^2$

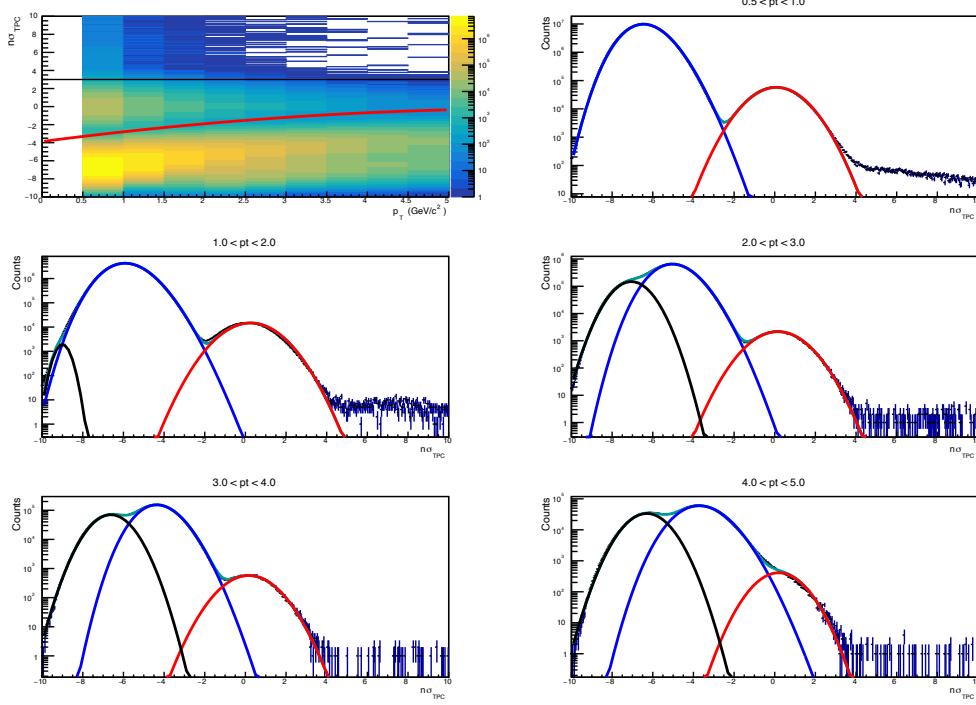


Figure 3.1: Top left panel: $n\sigma_{\text{TPC}}^{\text{ele}}$ as a function of electron p_T after applying the particle identification criteria on the TOF signal (see text for details). Remaining panels: the distribution of $n\sigma_{\text{TPC}}$ for different p_T^{ele} intervals.

shown for tracks within $|n\sigma_{\text{TOF}}^{\text{ele}}| < 3$. The applied criterion on the TPC dE/dx signal to select electron candidates is: $-3.9 + 1.17p_T - 0.094p_T^2 < |n\sigma_{\text{TPC}}^{\text{ele}}(p_T)| < 3$ (with p_T in units of GeV/c), which is represented by the red lines. The p_T -dependent lower limit on $|n\sigma_{\text{TPC}}^{\text{ele}}|$ is optimized to reject hadrons, following the same strategy as the previous analysis [21, 33]. The rest of the panels show the $n\sigma_{\text{TPC}}^{\text{ele}}$ distribution in different electron p_T intervals, projected from the top left panel of Fig. 3.1. A combination of two Gaussian functions (red for electrons, black for kaons and protons) and a Landau function (blue for pions) is used to fit the spectrum and isolate the various components. The comparison of the p_T intervals shows the increasing overlap as the transverse momentum gets higher.

The p_T -dependent lower limit on $|n\sigma_{\text{TPC}}^{\text{ele}}|$ is determined by studying the purity and hadron contamination of the electron candidates with varied $n\sigma_{\text{TPC}}^{\text{ele}}$ selections. The electron purity and hadron contamination are estimated by integrating the $n\sigma_{\text{TPC}}$ distribution, as shown in Eq. 3.1 and Eq. 3.2:

$$\text{Purity} = \frac{\int_{x_{\min}}^{x_{\max}} f_{\text{ele}}(x) dx}{\int_{x_{\min}}^{x_{\max}} (f_{\pi}(x) + f_{p/k}(x) + f_{\text{ele}}(x)) dx}, \quad (3.1)$$

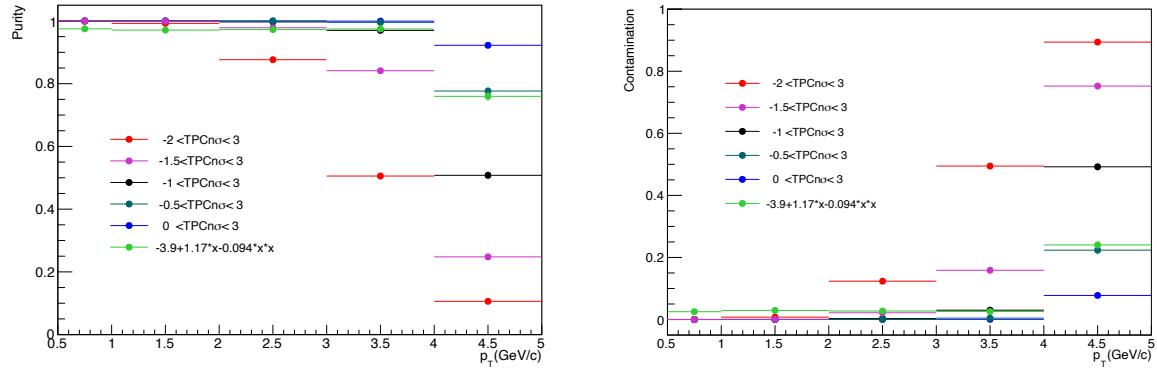


Figure 3.2: Left panel: the purity of electron with varied $n\sigma_{\text{TPC}}$ selections. Right panel: the contamination of hadrons with varied $n\sigma_{\text{TPC}}$ selections.

$$\text{Contamination} = \frac{\int_{x_{\min}}^{x_{\max}} (f_{\pi}(x) + f_{p/k}(x)) dx}{\int_{x_{\min}}^{x_{\max}} (f_{\pi}(x) + f_{p/k}(x) + f_{\text{ele}}(x)) dx}, \quad (3.2)$$

where x_{\min} and x_{\max} denote the lower and upper limits, respectively. An electron purity of 98% is achieved over the whole p_T range.

For those electrons from conversions in the material at large radii outside the innermost SPD layers are removed by the requirement of having associated hits in the two SPD layers out of the ITS [147, 148]. Electrons from conversion occurring in the beam pipe or inside the SPD material can pass the primary track selection criteria. Further rejection of background electrons originating from Dalitz decays of neutral mesons and photon conversions in the detector material (“photonic” electrons) is obtained by using a technique based on the invariant mass of the e^+e^- pairs [149, 150]. To maximize the probability of finding the electron partner from photonic decays, a prefilter selection is applied: the electron candidates are paired with opposite-sign tracks in the same event with loose identification ($|n\sigma_{\text{TPC}}^{\text{ele}}| < 5$ without any TOF requirements) and rejected if the invariant mass of the formed e^+e^- pair smaller than $50 \text{ MeV}/c^2$. The invariant mass distribution of the electron pairs in Fig. 3.3 shows a peak near $m_{ee} \sim 0 \text{ GeV}/c^2$, corresponding to the photonic decays, which is heavily suppressed by the prefilter selection.

3.2.2 Ξ reconstruction

The characteristics of Ξ^- are shown in Tab. 3.2. The reconstruction of Ξ^- baryons is exploited with the decay chain $\Xi^- \rightarrow \Lambda\pi^-$ ($\text{BR} = 99.887 \pm 0.035\%$), followed by $\Lambda \rightarrow p\pi^-$ ($\text{BR} = 63.9 \pm 0.5\%$) [1]. Tracks used to define Ξ^- candidates are required to have at least 70 associated crossed TPC pad rows, within 4σ of dE/dx for protons (pions). The Ξ^- and Λ baryons have long lifetimes ($c\tau$ of about 4.91

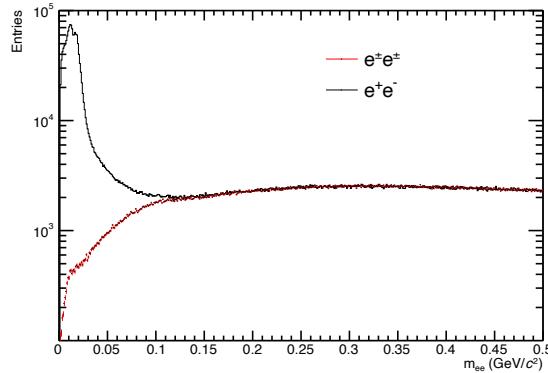


Figure 3.3: Invariant-mass distribution of electron pairs, the red one is with standard cuts and the black one is with prefiltered cuts.

cm and 7.89 cm, respectively [1]), therefore, they can be selected by using their different characteristic decay topologies [151]. The radial distances of the Ξ^- and Λ decay vertices from the beam lines, so-called decay length should be greater than 0.4 cm and 2.67 cm, respectively. The cosine of the pointing angle of the reconstructed Ξ^- to the primary vertex, and of the Λ to Ξ^- vertex, both required to be larger than 0.983. The DCA of V0 daughters, pions and protons, to the primary vertex should not be less than 0.073 cm. Pions originally decayed from Ξ^- are selected with the distance of the closest approach (d_0) of their tracks to the primary vertex greater than 0.05 cm. The selection criteria, as listed in Tab. 3.3, are tuned to reduce the background and enhance the purity of the signal. In the Fig. 3.5, the Ξ^- peak in the $\pi^- \Lambda$ invariant mass distribution integrated for $p_T^{\Xi^-} > 0$ is shown. Only Ξ^- candidates with invariant mass within 8 MeV/ c^2 from the world average Ξ^- mass (1321.71 ± 0.07 MeV/ c^2 [1]), indicated by an arrow in Fig. 3.5, are kept for further analysis.

Table 3.2: Properties of Ξ . The valence quark content, the strangeness S, the isospin, the spin and parity $I(J^P)$, the mass and the main decay channel, together with its branching ratio (BR) and decay length $c\tau$ are listed.

Particle	Quarks	S	$I(J^P)$	Mass (MeV/ c^2)	Decay particles	B.R. (%)	$c\tau$ (μm)
$\Xi^- (\bar{\Xi}^+)$	$dss(\overline{dss})$	-2 (+2)	$\frac{1}{2} (-\frac{1}{2})$	1321.71	$\Lambda\pi^- (\bar{\Lambda}\pi^+)$	99.9	4.91

3.2.3 $e\Xi$ pairs

The Ξ_c^0 candidates are reconstructed from $e^\pm \Xi^\mp$ pairs. Only pairs with an opening angle smaller than 90 degrees and the invariant mass $m_{e\Xi} < 2.5$ GeV/ c^2 are used for the analysis. Following the same approach in Ref. [34], the background contributions are estimated by exploiting the fact that Ξ_c^0 baryons and their antiparticles decay only into $e\Xi$ pairs with opposite charge sign, $e^\pm \Xi^\mp$ and $e^\mp \Xi^\pm$, called right-sign (RS), and not into same-sign pairs ($e^\pm \Xi^\pm$, denoted as wrong-sign (WS), and the assumption that

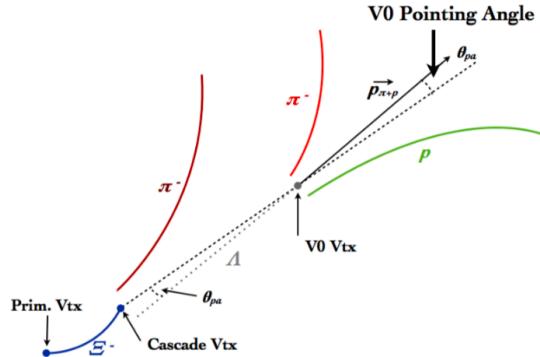


Figure 3.4: The topological variables are sketched out.

Table 3.3: The track and topology selections for Ξ applied in this analysis.

Cuts variables	cuts
Number of CrossedRows	> 70
CrossedRows Over Findable Cluster	> 0.77
Λ Mass tolerance (MeV/c^2)	7.5
Ξ Mass tolerance (MeV/c^2)	8
DCA of V0 to PV (cm)	> 0.05
DCA of V0 daughters to PV (cm)	> 0.073
V0 cosine of pointing angle to Ξ vertex	> 0.983
Ξ cosine of pointing angle to PV	> 0.983
DCA of bachelor track to PV (cm)	> 0.05
V0 decay length (cm)	> 2.67
Ξ decay length (cm)	> 0.4
$ n\sigma_{\text{TPC}} $ (proton)	< 4
$ n\sigma_{\text{TPC}} $ (pion)	< 4

combinatorial background candidates contribute equally to both RS and WS pairs. With this, the Ξ_c^0 raw yield can be obtained by subtracting the WS contribution from the RS pairs. Other contaminated contributions, contributing only to WS pairs, such as those from $\Xi_b^{0,-}$ semileptonic decays which do not give rise to RS pairs, will be corrected for after the subtraction, as described in Section 3.3.2.

In Fig. 3.6, the top left panel shows the invariant mass spectra of RS and WS $e\Xi$ pairs integrated over the whole p_T interval, $2 < p_T < 8 \text{ GeV}/c$. The Ξ_c^0 raw yield is obtained from the invariant mass distribution of RS pairs after subtracting the WS contribution shown in the top right panel. Due to the missing

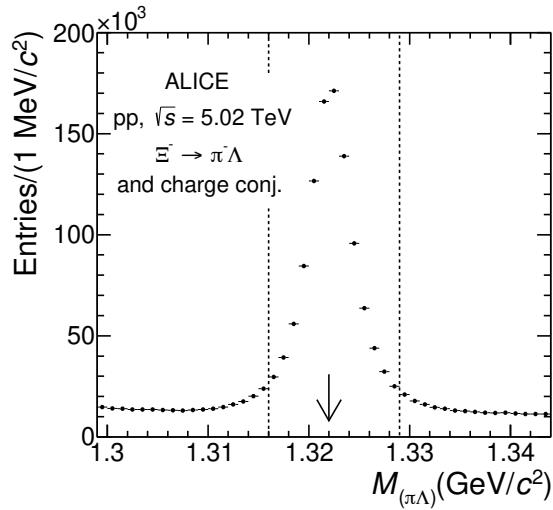


Figure 3.5: Invariant mass distribution of $\Xi^- \rightarrow \pi^- \Lambda$ (and charge conjugate) candidates integrated over $p_T^{\Xi^-}$. The arrow indicates the world average Ξ^- mass [1] and the dashed lines define the interval in which the Ξ^- candidates are selected for the Ξ_c^0 reconstruction (see text for details). Figure from [33].

momentum of neutrino, the invariant mass distribution of $e^+ \Xi^-$ pairs does not show a peak at the Ξ_c^0 mass. The p_T distribution of RS and WS $e\Xi$ pairs, where the range of invariant mass from 1.3 to 2.5 GeV/c^2 is integrated, seen in the bottom left panel. Then the ratio between the RS and WS spectra as a function of $p_T^{e\Xi}$ is also presented in the bottom right panel.

3.3 Corrections for raw yield extraction

The raw yield obtained by subtracting the WS pairs from RS pairs, (the distribution of the red and the black, as shown in the bottom left panel of Fig. 3.6), has to be corrected for three aspects: 1) the signal loss due to the mistagging of photonic electrons, called prefilter correction in Section 3.3.1; 2) the contaminated contribution from $\Xi_b^{0,-}$ to the WS pairs, which induces oversubtraction, explained in Section 3.3.2; 3) the missing neutrino momentum, which will be corrected by the Bayesian unfolding technique [152] implemented in the RooUnfold package [153], to obtain the Ξ_c^0 raw yield in intervals of $\Xi_c^0 p_T$, as shown in Section 3.3.3.

3.3.1 Prefilter correction

The probability of wrongly tagging an electron as photonic is estimated by applying the tagging algorithm, described in Section 3.2.1, to $e^\pm e^\pm$ pairs. The prefilter efficiency $\varepsilon_{\text{prefilter}}$ is given by Eq. 3.3:

$$\varepsilon_{\text{prefilter}} = \frac{N_{e\Xi}(\text{same-sign prefilter on})}{N_{e\Xi}(\text{prefilter off})} \quad (3.3)$$

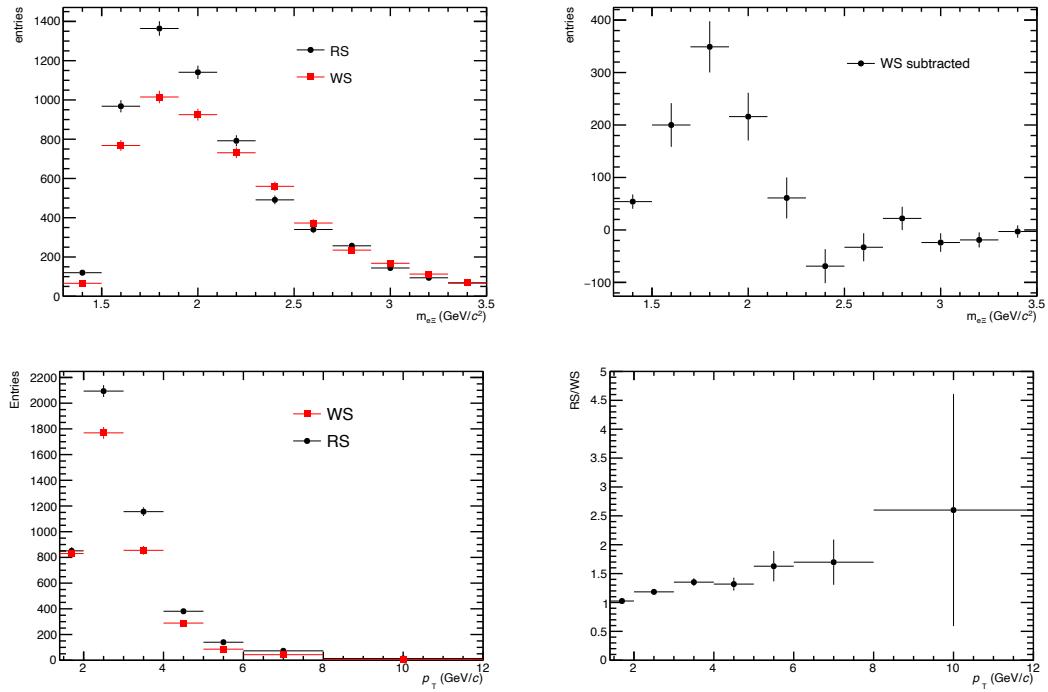


Figure 3.6: Top left panel: the invariant mass distributions of the RS and WS $e\Xi$ pairs. Top right panel: the WS subtracted spectra from RS, raw yield. Bottom left panel: the p_T distributions of the RS and WS $e\Xi$ pairs. Bottom right panel: the ratio between RS and WS spectra as a function of $p_T^{e\Xi}$.

where the $N_{e\Xi}$ (same-sign prefilter on) is the number of $e\Xi$ pairs having the same-sign for which the electron has a same-sign partner forming a pair with an invariant mass lower than the prefilter cut 50 MeV/ c^2 . $N_{e\Xi}$ (prefilter off) means the number of $e\Xi$ pairs for which the electron pairs without the prefilter mass cuts applied.

The calculated efficiency is shown in the left panel of Fig. 3.7. The yield comparison after the prefilter correction is in the right panel of Fig. 3.7. The red distribution is used for further analysis. The resulting correction is smaller than 2% for $p_T < 5 \text{ GeV}/c$, with a mild dependence on the p_T of the $e^+\Xi^-$ pair, which is also observed in Refs. [21, 34].

3.3.2 $\Xi_b^{0,-}$ oversubtraction correction

The decays of $\Xi_b^{0,-}, \Xi_b^{0,-} \rightarrow H_c e^- X$, where the H_c is any charmed baryon, followed by H_c decays to Ξ^- , $H_c \rightarrow \Xi^- X$, bring contaminated contribution to the WS invariant mass distribution, not to the RS one, resulting in a background over-subtraction. To estimate this contribution, assumptions must be made for the branching ratio of $\Xi_b^{0,-}$ into $e^- \Xi^- \bar{\nu}_e X$, as well as for the $\Xi_b^{0,-}$ production cross sections, since they are not measured yet.

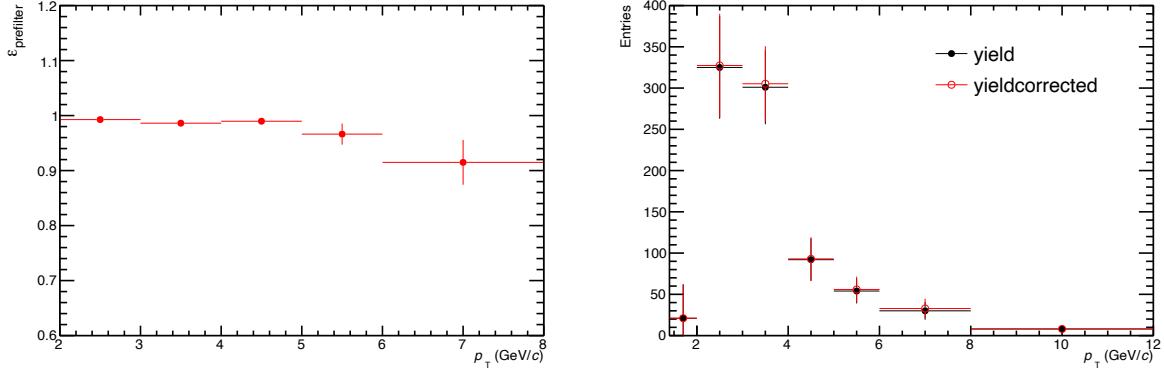


Figure 3.7: Left panel: Prefilter efficiency ($\epsilon_{e\Xi}$) in different intervals $p_T^{e\Xi}$. Right panel: The comparisons of yield distribution after prefilter correction.

Firstly, the shape of the transverse momentum distribution of $\Xi_b^{0,-}$ is assumed to be the same as the p_T distribution of Λ_b^0 baryons. The p_T -differential Λ_b^0 production cross section multiplied by the $\text{BR}(\Lambda_b^0 \rightarrow J/\psi \Lambda)$ in pp collisions at $\sqrt{s} = 7$ TeV has been measured by the CMS Collaboration [77]. The spectrum is fit to the Tsallis function [154]:

$$C \cdot p_T \cdot \left[1 + \frac{\sqrt{p_T^2 + m^2} - m}{nT} \right]^{-n}, \quad (3.4)$$

where C is a normalisation parameter, T and n are shape parameters, m is the mass of the b hadron. The parameters used are: $T = 1.1$ GeV/c and $n = 7.6 \pm 0.4$, respectively. The Tsallis function is assumed valid down to $p_T = 0$, as the green distribution shown in the right panel of Fig. 3.8. The distributions of the beauty-hadron cross sections at $\sqrt{s} = 7$ TeV and 5.02 TeV obtained with FONLL pQCD calculations [3, 155] are shown in the left panel of Fig. 3.8. The ratio between them is used to obtain the Λ_b^0 measurement at the centre-of-mass energy of 5.02 TeV, which is the black distribution shown in the right panel of Fig. 3.8.

The second assumption is that the fragmentation fraction of beauty quarks hadronising into Λ_b^0 and $\Xi_b^{0,-}$ is the same as those in e^+e^- collisions. The yield of $\Xi_b^{0,-} \rightarrow e^-\Xi^-\bar{\nu}_e X$ is therefore computed using: i) the \sqrt{s} -scaled Λ_b^0 cross section at $\sqrt{s} = 5.02$ TeV, shown in the right panel of Fig. 3.8, ii) the values of $f(b \rightarrow \Xi_b^{0,-}) \times \text{BR}(\Xi_b^{0,-} \rightarrow e^-\Xi^-\bar{\nu}_e X)$ [1] and $f(b \rightarrow \Lambda_b^0) \times \text{BR}(\Lambda_b^0 \rightarrow J/\psi \Lambda)$ [156] obtained from e^+e^- collisions, iii) the $\Xi_b^{0,-} \rightarrow e^-\Xi^-\bar{\nu}_e X$ acceptance \times efficiency (Acc \times ϵ) from the simulations, shown in left panel of Fig. 3.9. The raw yield of Ξ_c^0 from decays of $\Xi_b^{0,-}$, is obtained by Eq. 3.5,

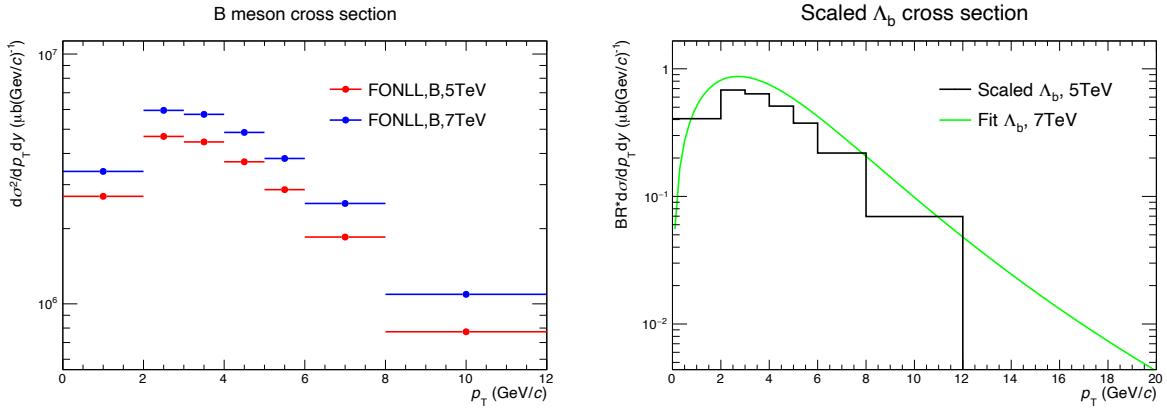


Figure 3.8: Left panel: the distributions of the beauty-hadron cross sections at $\sqrt{s} = 7 \text{ TeV}$ and 5.02 TeV obtained with FONLL pQCD calculations. Right panel: the fitted Λ_b^0 distribution at $\sqrt{s} = 7 \text{ TeV}$, and the scaled Λ_b^0 distribution at $\sqrt{s} = 5.02 \text{ TeV}$.

$$N_{\Xi_b}^{\text{raw}} = \frac{d\sigma_{\Xi_b}^2}{dp_T dy} \cdot 2 \cdot \Delta p_T \cdot \Delta y \cdot \text{Acc} \times \varepsilon \cdot \mathcal{L}_{\text{int}}, \quad (3.5)$$

It will be used for computing the counts of $e\Xi$ pairs. The spectrum of $e\Xi$ pairs coming from $\Xi_b^{0,-}$ can be estimated with the refolded procedure (discussed in the next section 3.3.3), applied with the response matrix of $\Xi_b^{0,-}$ and $e\Xi$ in the right panel of Fig. 3.9.

The comparison of p_T distribution of $e\Xi$ pairs between from the default analysis (RS-WS) (black) and from the contribution of $\Xi_b^{0,-}$ (red) is illustrated in the top left panel of Fig. 3.10, their ratio is shown in the top right panel Fig. 3.10. The $e\Xi$ pairs from the bottom baryon are added to the raw yield spectra (RS-WS), the comparison of the final corrected yield of $p_T^{\Xi_c^0}$ is shown in the bottom left panel. The final result increases by about 2% after correcting the over-subtraction from $\Xi_b^{0,-}$, as seen in the right panel.

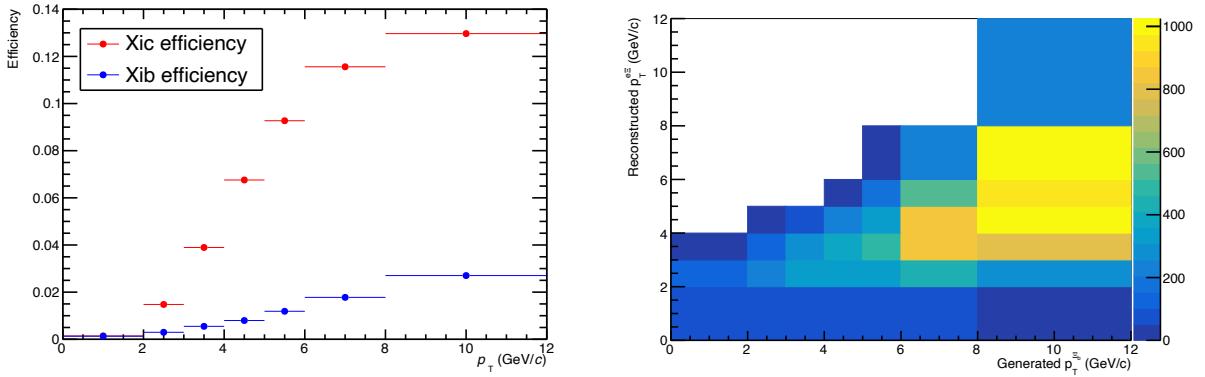


Figure 3.9: Left panel: the acceptance-times-efficiency of $\Xi_b^{0,-}$. Right panel: the response matrix of $\Xi_b^{0,-}$ and $e\Xi$.

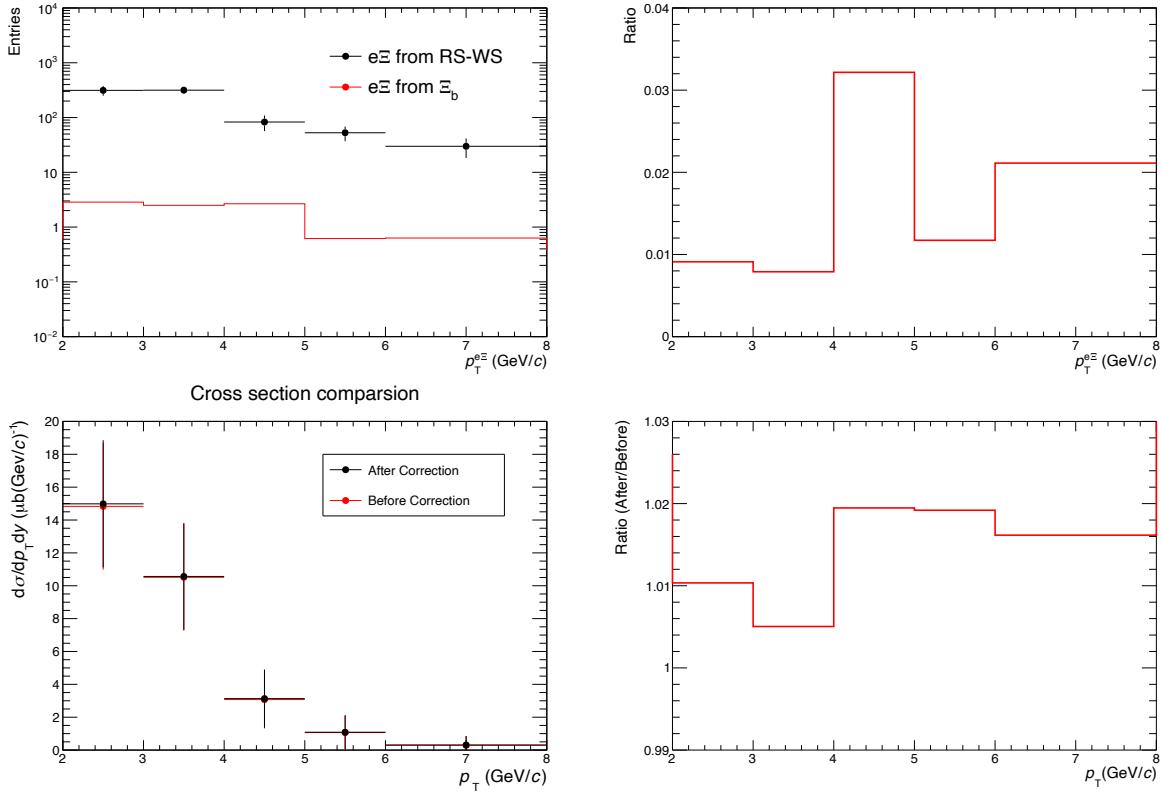


Figure 3.10: Top left panel: the comparison of distribution of $p_T^{e\Xi}$ from default analysis (black) and from $\Xi_b^{0,-}$ (red). Top right panel: the ratio between red and black, which represents the amount of over-subtraction. Bottom left panel: the comparison of the final result between the WS subtracted spectra without over-subtraction correction (red), and with over-subtraction correction (black). Bottom right panel: the ratio between the corrected and uncorrected result.

After the two corrections discussed above in Section 3.3.2 and Section 3.3.1, the invariant mass distributions of WS and RS pairs in the interval $2 < p_T^{e\Xi} < 8$ GeV/c are shown in the left panel of Fig. 3.11, and in the right panel is the invariant mass distribution of Ξ_c^0 candidates obtained after subtracting the WS pair yield from the RS yield. To correct for the missing momentum of the neutrino, the unfolding technique is used, converting $p_T^{e\Xi}$ into Ξ_c^0 , which is discussed in next section.

3.3.3 Unfolding procedure

The detector response, described by a matrix R , represents the correlation between the p_T of the Ξ_c^0 baryon (true distribution T) and that of the reconstructed $e^+\Xi^-$ pair (measured distribution M). The response matrix is determined via a simulation with the PYTHIA 8.243 event generator [13] and the GEANT 3 transport code [157], including a realistic description of the detector conditions and alignment during the data taking period. Mathematically, matrix multiplication can be used to express this relation: a true p_T distribution T is folded with the detector response R to the measured p_T distribution M , as

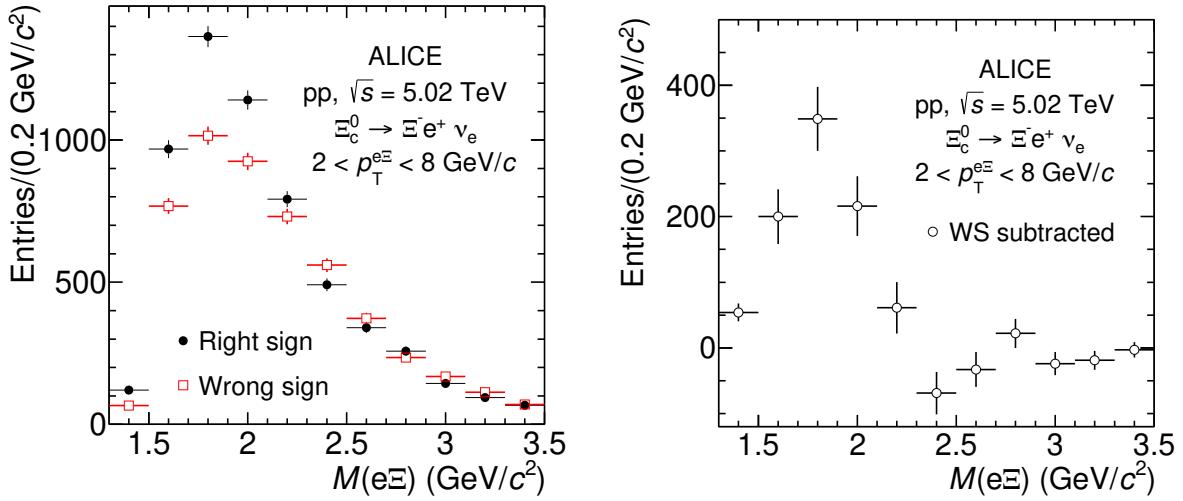


Figure 3.11: Left panel: invariant mass distributions of right-sign and wrong-sign $e\Xi$ pairs with $2 < p_T < 8$ GeV/ c . Right panel: invariant mass distribution of Ξ_c^0 candidates obtained by subtracting the wrong-sign pair yield from the right-sign pair yield. Figures from [33].

Eq. 3.6. To determine the true p_T distribution T , the measured p_T distribution M needs to be *unfolded* by R^{-1} , Eq. 3.7:

$$M = T \times R, \quad (3.6)$$

$$T = (R^{-1}) \times M, \quad (3.7)$$

In fact, the inversion R^{-1} would give an unreliable result with very large variances. This problem can be overcome by turning it into an iterative procedure. Bayesian unfolding is the default method employed in this analysis, which is an alternative approach to unfold the measured distribution based on Bayes' theorem to the true distribution. The process is repeated with the previous unfolded distribution as a new input for the unfolding procedure until it converges.

Bayes' theorem describes the probability of an event, which is based on prior conditions, illustrated as in Eq. 3.8:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3.8)$$



$$\tilde{R}_{ji} = \frac{R_{ij}U_j}{\sum_k R_{ik}U_k}, \quad (3.9)$$

where A represents the true measurement T and B means the measurement M with the detector. The $P(A)$ and $P(B)$ are the prior probabilities for the true distribution T and the measured distribution M , respectively. $P(B)$ is calculated with the response matrix: $M_i = \sum_k R_{ik}U_k$. U_j is the prior distribution of the true distribution. $P(A)$ is defined by the U_j . $P(B|A)$ is the probability of the measured distribution M under the condition of the true distribution T and it is same with the matrix elements R_{ij} of the detector response matrix R . Analogously, $P(A|B)$ is the probability of the true distribution T under the condition of the measured distribution M , and it can be analogous to \tilde{R}_{ji} . In this analysis, the measured spectra M have been used as the prior distribution. In the above formulas Eq. 3.8 and Eq. 3.9, $P(A|B)$ and $P(A)$ are unknown. An iterative method is required to solve the equation:

$$U' = \tilde{R}[U] \cdot M, \quad (3.10)$$

If the spectra U is the true distribution T , the unfolded spectra U' is T . For the next iteration, U is replaced with U' to get the next solution U'' , and so on.

The response matrix needs to be determined by using a realistic Ξ_c^0 -baryon p_T distribution which is not known prior. Therefore, a two-step iterative procedure is adopted: 1) the response matrix is obtained from the MC simulation (p_T distribution generated with PYTHIA 8 [13]. This response matrix is used to calculate a first estimate of the Ξ_c^0 p_T -differential spectrum based on the measured p_T distribution of $e^+\Xi^-$ pairs. 2) The Ξ_c^0 p_T distribution obtained from this first iteration is used for the reweighting procedure, which is applied to the response matrix, then the weighted matrix is used for the second iteration. The weighting factor is calculated as the ratio of p_T distribution of the first estimation of Ξ_c^0 p_T (in blue) in data and of the MC generated Ξ_c^0 (in orange) as shown in the left panel of Fig. 3.12. In the right panel of Fig. 3.12, the obtained ratio is fitted with exponential functions in $p_T > 2 \text{ GeV}/c$.

The weighted response matrix is shown in the left panel of Fig. 3.13, which is defined in the transverse momentum interval $1.4 < p_T < 12 \text{ GeV}/c$. It is wider than the p_T interval used for the cross section measurement, to avoid possible edge effects at the lowest and highest p_T intervals of the measurement. In this analysis, the Bayesian procedure required three iterations to converge. The measured spectrum is compared with the unfolded spectra implemented with the *weighted* and *unweighted* response matrices. To check the RooUnfold stability, the refolded procedure is required. In principle, the refolded spectrum

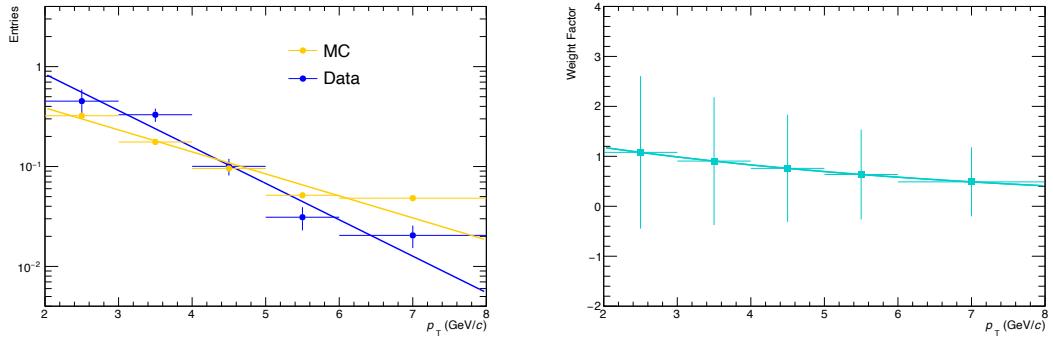


Figure 3.12: Left panel: the p_T distribution of Ξ_c^0 in data and in MC, fitted by the expol0 function. Right panel: the weighting factor is defined as the ratio of p_T distribution of Ξ_c^0 in data and the in MC.

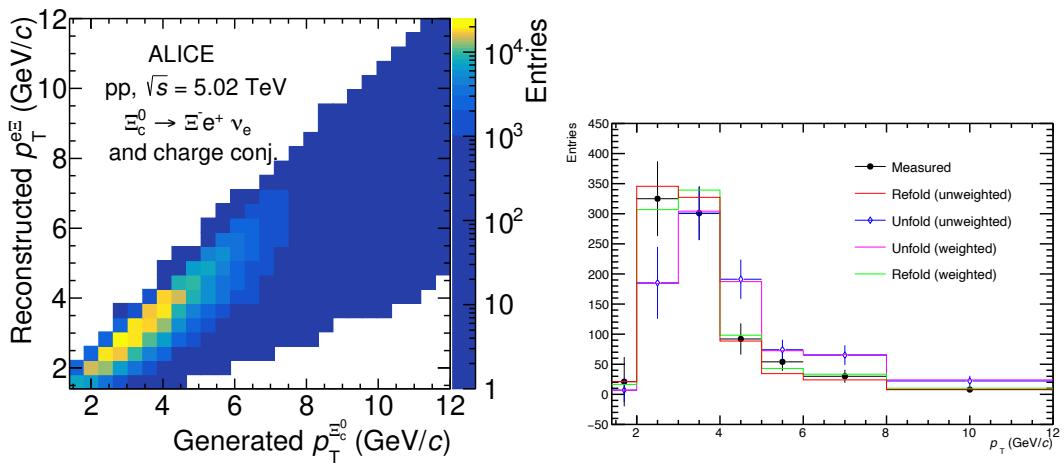


Figure 3.13: Left panel: Correlation matrix between the generated Ξ_c^0 -baryon p_T and the reconstructed $e^+\Xi^-$ pair p_T , obtained from the simulation based on PYTHIA 8 described in the text. Figure from [33]. Right panel: Comparisons of the p_T distribution of measured ($p_T^{e\Xi}$), unfolded ($p_T^{\Xi_c^0}$), and refolded ($p_T^{e\Xi}$).

is supposed to be matched with the measured one, as in Eq. 3.6. All the yield distributions can be seen in the right panel of Fig. 3.13. The comparison of the two refolded spectra shows that the green refolded one is closer to the measured distribution than the red one without weights applied, indicating that the reweighting procedure indeed helps in correcting the p_T shape of Ξ_c^0 .



3.4 Reconstruction efficiency and feed-down subtraction

The p_T -differential cross section of prompt Ξ_c^0 -baryon production is obtained as Eq. 4.21:

$$\frac{d^2\sigma_{\Xi_c^0}}{dp_T dy} = \frac{1}{BR} \cdot \frac{1}{2\Delta y \Delta p_T} \cdot \frac{f_{\text{prompt}} \cdot N_{\text{raw}}^{\Xi_c^0}}{(Acc \times \varepsilon)_{\text{prompt}}} \cdot \frac{1}{\mathcal{L}_{\text{int}}}, \quad (3.11)$$

where $N_{\text{raw}}^{\Xi_c^0}$ is the raw yield after those corrections mentioned in Section 3.3 in a given p_T interval with width Δp_T , as pink one in the right panel of Fig. 3.13. f_{prompt} represents the fraction of prompt Ξ_c^0 in the raw yield, and BR is the branching ratio of Ω_c^0 via its semileptonic decay mode $e^+ \Omega^- \nu_e$, and \mathcal{L}_{int} is the integrated luminosity. The $(Acc \times \varepsilon)_{\text{prompt}}$ factor means the product of detector acceptance and efficiency for prompt Ξ_c^0 baryons, in which the ε is the reconstruction and selection of the Ξ_c^0 decay-product tracks and the Ξ_c^0 -candidate selection. The factor Δy is the width of the rapidity interval in which the generated Ξ_c^0 are considered, and is used to calculate the cross section in one unit of rapidity. The factor 1/2 counts that $N_{\text{raw}}^{\Xi_c^0}$ includes both particles and antiparticles, therefore, the cross section is given only for particles. The results of the ratio $BR(\Xi_c^0 \rightarrow \Xi^- e^+ \nu_e) / BR(\Xi_c^0 \rightarrow \Xi^- \pi^+) = 1.36 \pm 0.14$ (stat.) ± 0.19 (syst.) measured by ALICE in pp collisions at $\sqrt{s} = 13$ TeV [21] and the measured hadronic decay branching ratio $BR(\Xi_c^0 \rightarrow \Xi^- \pi^+)$ reported in the PDG [1], bring the value of $BR(\Xi_c^0 \rightarrow \Xi^- e^+ \nu_e)$ to be $(1.94 \pm 0.55)\%$.

The $(Acc \times \varepsilon)$ factor is obtained from the same simulations, which are also used to determine the response matrix. The $(Acc \times \varepsilon)$ is calculated separately for prompt and feed-down Ξ_c^0 baryons, which is reported in the left panel of Fig. 3.14. It shows that the efficiencies of prompt and feed-down baryons are consistent with each other within uncertainties, as the applied selection criteria are not sensitive to the displacement by a few hundred micrometers of the prompt and feed-down Ξ_c^0 decay vertices from the collision point. To compute the efficiency with a realistic momentum distribution of Ξ_c^0 baryons, the reweighted procedure is very important, aiming to make the measured one match better the p_T shape of the generated Ξ_c^0 baryons from the PYTHIA 8 simulation. The reweighted procedure is adopted with two iterative steps similar to the one implemented in the response matrix, explained as in Section 3.3.3.

To obtain the prompt production of Ξ_c^0 baryons, the factor f_{prompt} has to be calculated, which is:

$$f_{\text{prompt}} = 1 - \frac{N_{\text{raw}}^{\Xi_c^0 \text{ feed-down}}}{N_{\text{raw}}^{\Xi_c^0}} = 1 - \frac{(Acc \times \varepsilon)_{\text{feed-down}} \cdot \Delta y \cdot \Delta p_T \cdot BR \cdot \mathcal{L}_{\text{int}}}{N_{\text{raw}}^{\Xi_c^0 + \Xi_c^0} / 2} \times \left(\frac{d^2\sigma}{dp_T dy} \right)_{\Xi_c^0 \text{ feed-down}}, \quad (3.12)$$

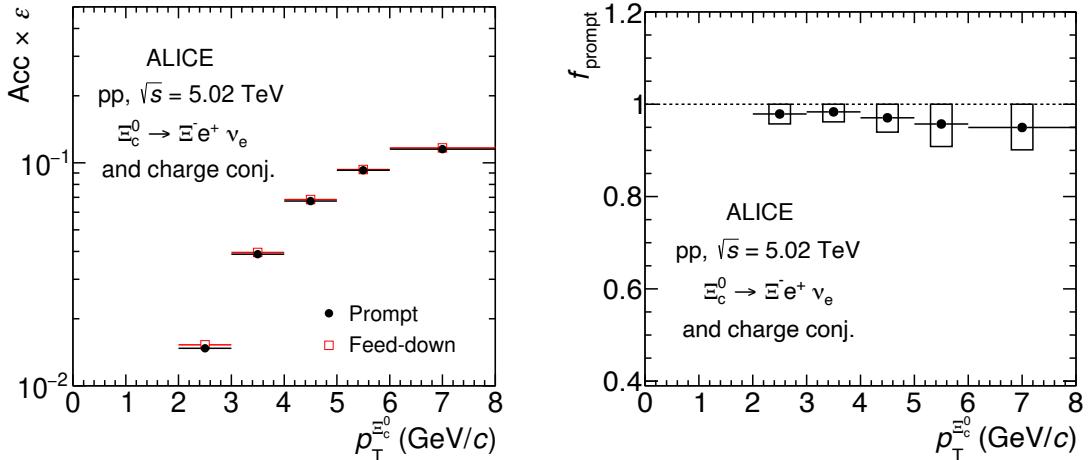


Figure 3.14: Left panel: product of acceptance and efficiency for prompt and feed-down Ξ_c^0 baryons in pp collisions at $\sqrt{s} = 5.02$ TeV as a function of p_T . Right panel: fraction of prompt Ξ_c^0 baryons in the raw yield (f_{prompt}) as a function of p_T . The systematic uncertainties of f_{prompt} are shown as boxes (explained in Section 3.5.7). Figures from [33].

where $N_{\text{raw}}^{\Xi_c^0 + \bar{\Xi}_c^0}/2$ means the averaged raw yields of particles and antiparticles, $(\text{Acc} \times \epsilon)_{\text{feed-down}}$ is the product of detector acceptance and efficiency for feed-down Ξ_c^0 baryons, as the red distribution in the left panel of Fig. 3.14. The p_T -differential cross section of feed-down Ξ_c^0 baryon production is denoted as $\left(\frac{d^2\sigma}{dp_T dy}\right)_{\Xi_c^0 \text{ feed-down}}$. However, the production cross section of Ξ_c^0 from beauty-baryon decays is not known yet, an alternative method based on the measured cross section of feed-down Λ_c^+ in Ref. [11] is adopted. The production of Λ_c^+ from Λ_b^0 -baryon decays is estimated using the b-quark p_T -differential cross section from FONLL calculations, multiplied by the beauty quarks fraction that fragment into Λ_b^0 . The beauty fragmentation fraction in pp collisions at $\sqrt{s} = 13$ TeV [76], has been measured by the LHCb Collaboration, taking into account its p_T dependence. The $\Lambda_b^0 \rightarrow \Lambda_c^+ + X$ decay kinematics is modeled using PYTHIA 8.243 simulations [13]. The cross section of Λ_c^+ decayed from Λ_b^0 is scaled by the fraction of Ξ_b^- decaying in a final state with Ξ_c^0 over the fraction of Λ_b^0 decaying into Λ_c^+ . Their fractions are taken to be 50% and 82%, respectively, obtained from the PYTHIA 8.243 generator [13].

The cross section of Ξ_c^0 from beauty feed-down is calculated from the one of Λ_c^+ originating from Λ_b^0 decays, and scaled by the ratio of the measured p_T -differential yields of inclusive Ξ_c^0 and prompt Λ_c^+ baryons, as reported in Fig. 3.25. Two assumptions are required: i) the p_T shape of the cross sections of feed-down Ξ_c^0 is the same as the one of feed-down Λ_c^+ ; ii) the ratio of inclusive Ξ_c^0/Λ_c^+ is the same as the ratio for the feed-down case. This procedure is reliable as the inclusive Λ_c^+ -baryon yield is dominated by the prompt production, based on the f_{prompt} values close to unity reported in Ref. [10]. More discussion about the systematic uncertainty of this procedure can be found in Section 3.5.7.



博士学位论文
DOCTORAL DISSERTATION



The value of f_{prompt} as a function of p_T is shown in the right panel of Fig. 3.14. The vertical bars and empty boxes represent the statistical and systematic uncertainties (explained later in Section 3.5.7).



3.5 Systematic uncertainties

The systematic uncertainty on the Ξ_c^0 production cross section has different contributions, which will be discussed in the following.

In order to prevent an effect of systematic uncertainty from statistical fluctuation, a Barlow value [158] for variations is calculated based on:

$$B = \frac{|\text{Default} - \text{Variation}|}{\sqrt{\sigma_{\text{Default}}^2 - \sigma_{\text{Variation}}^2}}. \quad (3.13)$$

In Eq. 3.13, if the Barlow value (B) is less than 1, differences could be from a statistical fluctuation, if greater than 1, could result in a systematic uncertainty.

3.5.1 Track efficiency

The systematic uncertainty on the track efficiency in this analysis consists of three resources: i) by varying the TPC track selection criteria for electron tracks in Section 3.5.1.1, ii) by varying the TPC track selection criteria also for the Ξ^- daughter tracks in Section 3.5.1.2, iii) and by comparing the probability of prolonging a track from the TPC to ITS (“matching efficiency” discussed previously in Section 2.3) in data and simulation, studied in Section 3.5.1.3.

3.5.1.1 Electron selection

The number of TPC clusters and the number of ITS clusters are varied to study the uncertainties on the electron selection. The variations for the systematic study are shown in Tab. 3.4.

Table 3.4: Varied criteria for electron selections. Selections that are not shown in this table are the same as the ones in Tab. 3.1.

Cut variables	VeryLoose	Loose	Standard	Tight	VeryTight
Number of crossed rows	> 65	> 65	> 70	> 75	> 80
The ratio of crossed rows and findable clusters	> 0.75	> 0.8	> 0.8	> 0.85	> 0.9
Number of TPC PID cluster	> 40	> 45	> 50	> 55	> 60

In Fig. 3.15, the comparison of the Ξ_c^0 spectrum with different selections and the ratios of the Ξ_c^0 spectra with variations to the standard cut, are reported in the top two panels. The root mean square (RMS) and Barlow values are calculated for each p_T bin. The systematic uncertainty obtained from RMS is assigned 2% for the electron track selection.

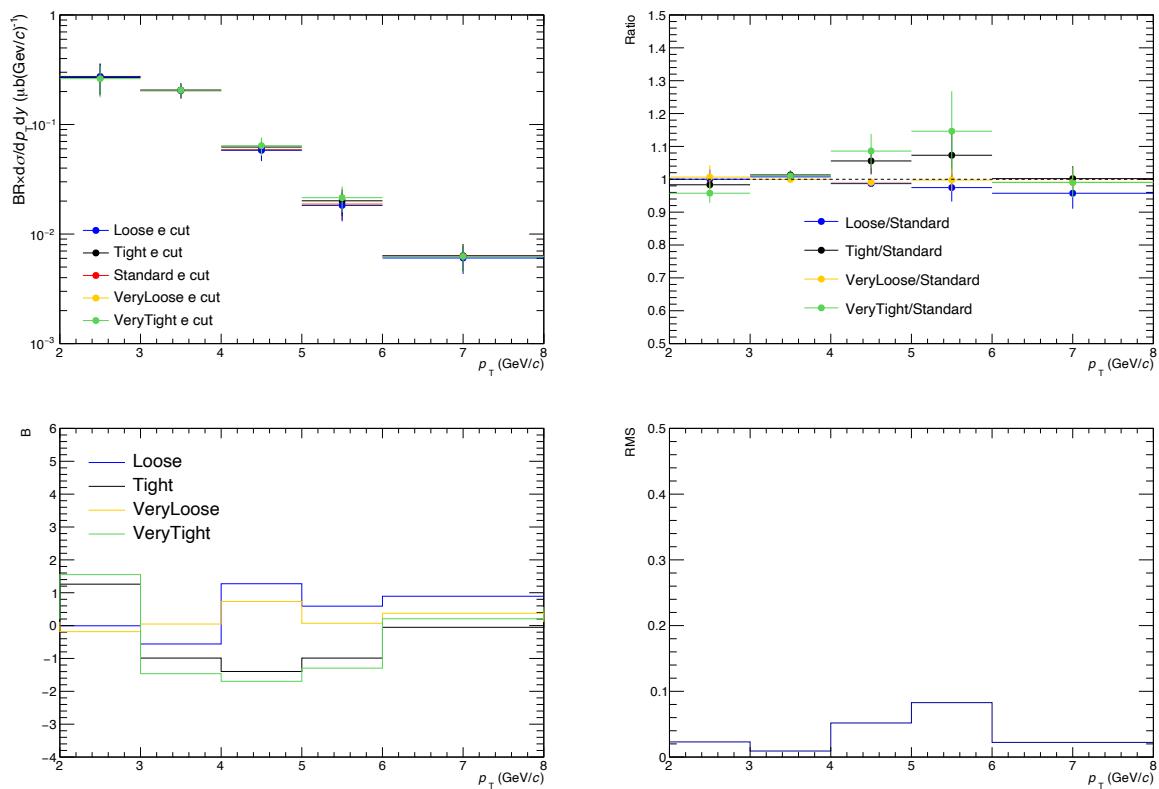


Figure 3.15: Top left panel: the Ξ_c^0 spectrum with different selections. Top right panel: ratios of the Ξ_c^0 spectrum with variations. Bottom left panel: Barlow values. Bottom right panel: the root mean square (RMS) of the deviations from 1.



3.5.1.2 Ξ reconstruction

The varied criteria associated with different numbers of TPC pad rows, shown in Tab. 3.5, is used to study the uncertainty on the track reconstruction of Ξ .

Fig. 3.16 shows the comparison of results with different selections. The systematic uncertainty is 4% for the Ξ^- daughters independent of $\Xi_c^0 p_T$.

Table 3.5: Ξ track cuts used to study the uncertainties in Ξ reconstruction. Other cuts are not shown which are the same as in Tab. 3.3.

Cuts variables	VeryLoose	Loose	Standard	Tight	VeryTight
Number of crossed rows	> 65	> 65	> 70	> 75	> 80
The ratio of crossed rows and findable clusters	> 0.70	> 0.75	> 0.77	> 0.79	> 0.81

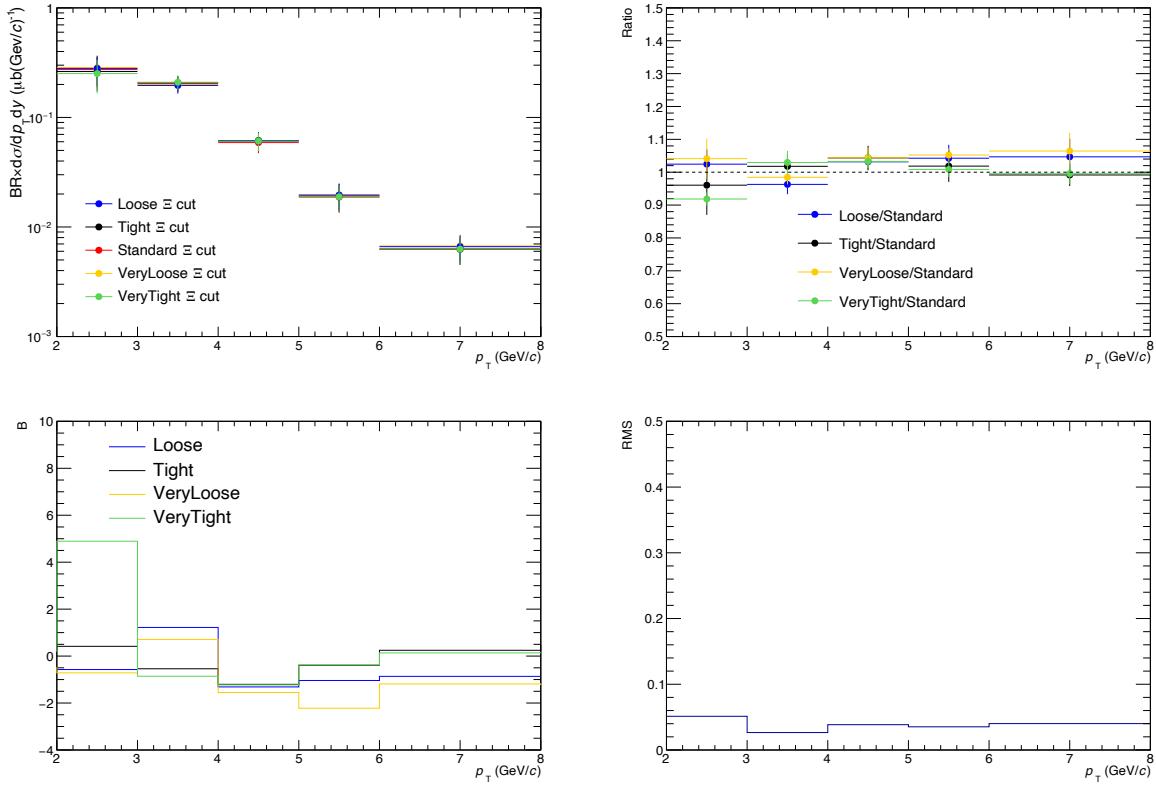


Figure 3.16: Top left panel: the Ξ_c^0 spectrum with different selections. Top right panel: ratios of the Ξ_c^0 spectrum with variations. Bottom left panel: Barlow value. Bottom right panel: the root mean square (RMS) of the deviations from 1.



3.5.1.3 ITS-TPC matching

The uncertainty on the matching efficiency affects only the electron track, and not the tracks of the Ξ^- decay particles, as its prolongation to ITS is not required. It is defined as the relative difference in the ITS-TPC matching efficiency between simulation and data. The uncertainty is propagated from the electron tracks to the Ξ_c^0 taking into account its decay kinematics, and it is assigned 2%, independent of $\Xi_c^0 p_T$, illustrated in Tab. 3.6. Details can be found in Appendix A.1.1.

Table 3.6: The systematic uncertainty of $p_T^{\Xi_c^0}$ propagated from $p_T^{e\Xi}$.

	Ξ_c^0				
p_T bins (MeV/c)	2-3	3-4	4-5	5-6	6-8
Systematic	2.05%	2.05%	2.06%	2.13%	2.2%

3.5.2 Electron identification

Systematic uncertainties can also arise from discrepancies in the particle-identification efficiency in data compared to that in MC simulation. The analysis is repeated by varying the selection criteria of the identification for the electron candidate tracks, shown in Tab. 3.7. These PID selection criteria are depicted in Fig. A.3 in Appendix A.1.2.

Fig. 3.17 shows the final results with different selection criteria. The systematic uncertainty is dependent on the $\Xi_c^0 p_T$, ranging from 4% to 7%.

3.5.3 Ξ topology

The varied criteria of the Ξ^- topology cuts are shown in Tab. 3.8. The systematic uncertainty is 6%, which is estimated from the RMS of the distribution of the Ξ_c^0 corrected yields, as reported in Fig. 3.18.

3.5.4 $e\Xi$ pair selection

The uncertainty of the $e^+\Xi^-$ -pair selection efficiency is estimated by varying the selection criteria for the invariant masses of $e\Xi$ pairs, seen in Tab. 3.9.

The systematic uncertainty is assigned for 3%, obtained by fitting the ratio with Pol0 function, seen in the right panel of Fig. 3.19.

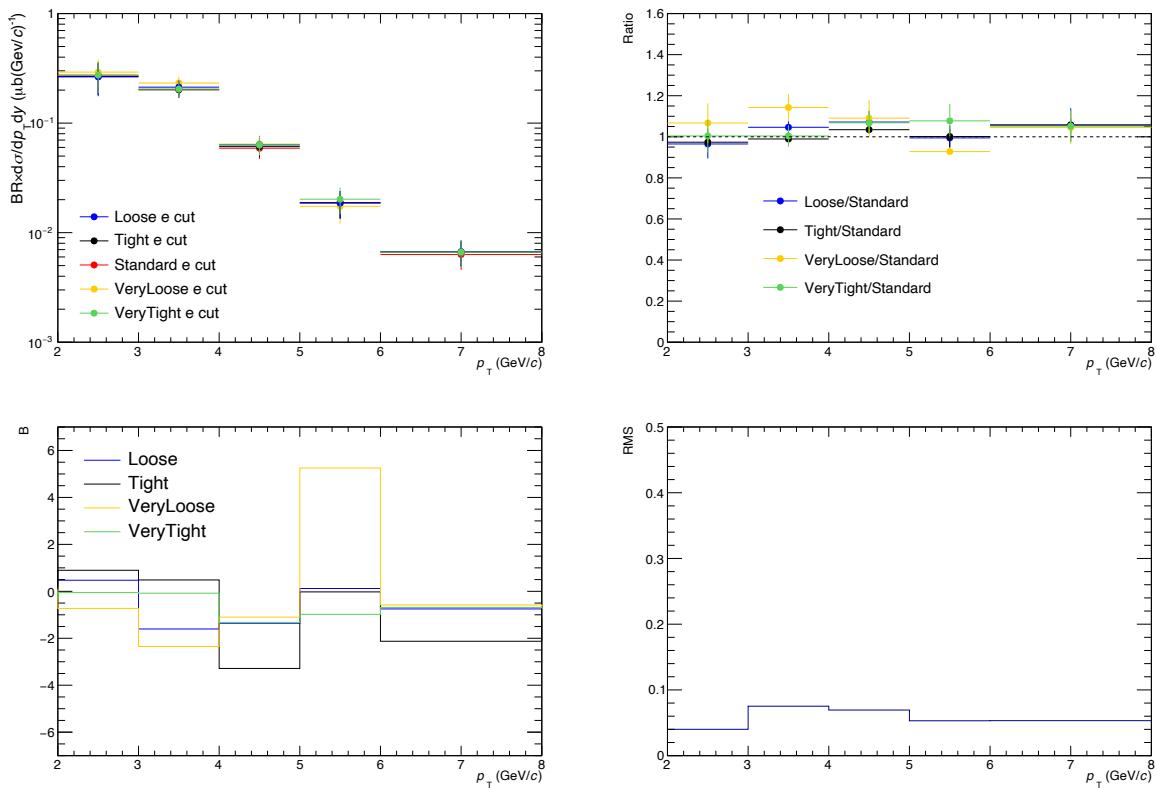


Figure 3.17: The systematic study for electron candidates identification. Top left panel: the Ξ_c^0 spectrum with different selections. Top right panel: ratios of the Ξ_c^0 spectrum with variations. Bottom left panel: Barlow value. Bottom right panel: the root mean square (RMS) of the deviations from 1.

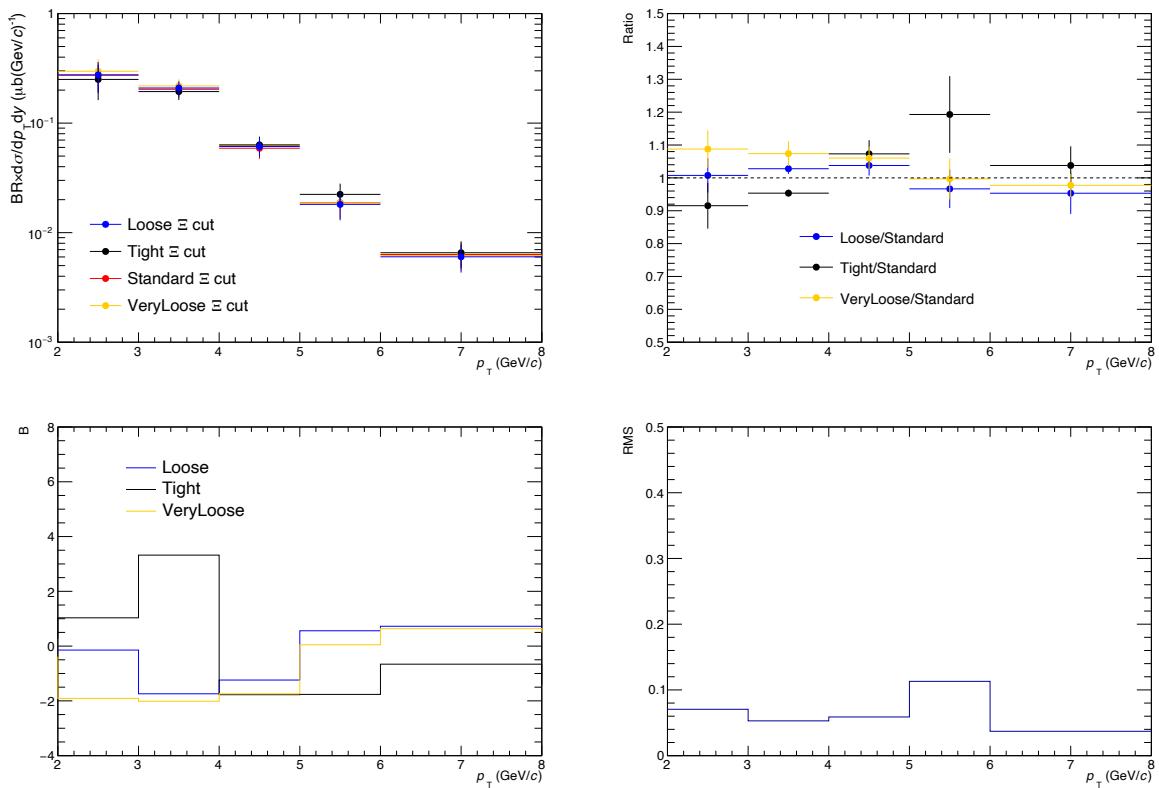


Figure 3.18: The systematic study for Ξ reconstruction. Top left panel: the Ξ_c^0 spectrum with different selections. Top right panel: ratios of the Ξ_c^0 spectrum with variations. Bottom left panel: barlow value. Bottom right panel: the root mean square (RMS) of the deviations from 1.



Table 3.7: Electron selections used to study the uncertainties of track identification. Cuts that are not shown in this table are the same as the ones in Tab. 3.1.

Cuts variables	Lower limit of $n\sigma_{\text{TPC}}^{\text{ele}}$	$n\sigma_{\text{TOF}}^{\text{ele}}$
VeryLoose	$-4.4 + 1.23p_T - 0.102p_T^2$	< 3
Loose	$-4.17 + 1.20p_T - 0.098p_T^2$	< 3
Standard	$-3.9 + 1.17p_T - 0.094p_T^2$	< 3
Tight	$-3.9 + 1.28p_T - 0.094p_T^2$	< 3
VeryTight	$-3.9 + 1.28p_T - 0.094p_T^2$	< 2

Table 3.8: The varied selections of Ξ topology to study the uncertainties for the daughter track reconstruction. Selections that are not shown in this table are the same as Tab. 3.3.

Cut variables	VeryLoose	Loose	Standard	Tight
Λ Mass tolerance (MeV/c^2)	7.5	7.5	7.5	7.5
Ξ Mass tolerance (MeV/c^2)	8	8	8	8
DCA of V0 to PV (cm)	> 0.01	> 0.02	> 0.03	> 0.04
DCA of V0 daughters to PV (cm)	> 0.05	> 0.061	> 0.073	> 0.088
V0 cosine of pointing angle to Ξ vertex	> 0.98	> 0.981	> 0.983	> 0.9839
DCA of bachelor track to PV (cm)	> 0.01	> 0.0146	> 0.0204	> 0.0275
V0 decay length (cm)	> 0.20	> 1.55	> 2.67	> 3.6
Ξ decay length (cm)	> 0.20	> 0.29	> 0.38	> 0.53

Table 3.9: The variation of pairs mass selection.

Cut variables	Loose	Standard
Invariant mass cut (GeV/c^2)	2.7	2.5

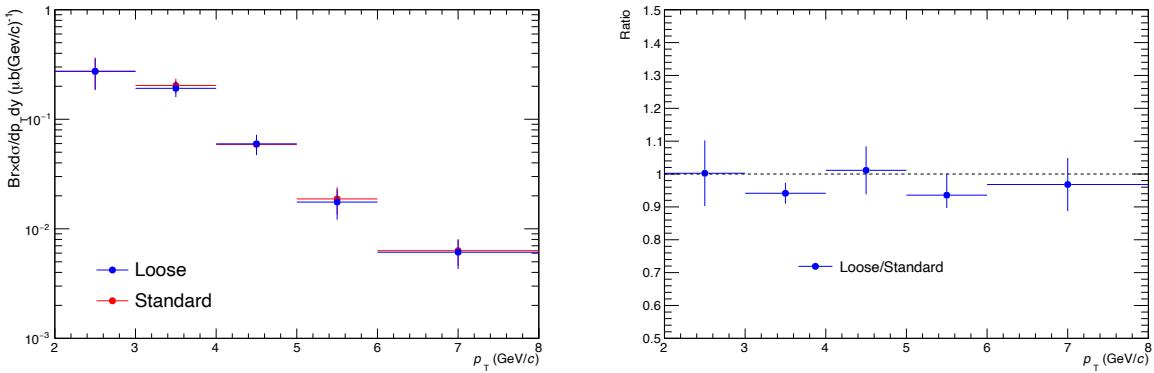


Figure 3.19: The systematic study for $e\Xi$ pair selection. Left panel: the Ξ_c^0 spectra with different selections. Right panel: ratios of the Ξ_c^0 spectra with variations.



3.5.5 Unfolding procedure

The systematic uncertainty of the correction for the missing neutrino momentum is studied by varying the unfolding procedure: 1) the number of iterations in the Bayesian unfolding procedure is varied in Section 3.5.5.1; 2) use a different unfolding method in Section 3.5.5.2, and 3) the p_T range and the binning of the response matrix used in the unfolding, seen in Section 3.5.5.3.

3.5.5.1 Bayesian iteration number

The iteration number 3 is used as default in the Bayesian unfolding. To study the systematic uncertainty of using different numbers, the iteration number is varied from 2 to 6.

The comparison of unfolded yield with different iteration numbers is seen in Fig. 3.20. The systematic uncertainty is assigned to be the RMS values as shown in the bottom panel of Fig. 3.20, dependent on $\Xi_c^0 p_T$.

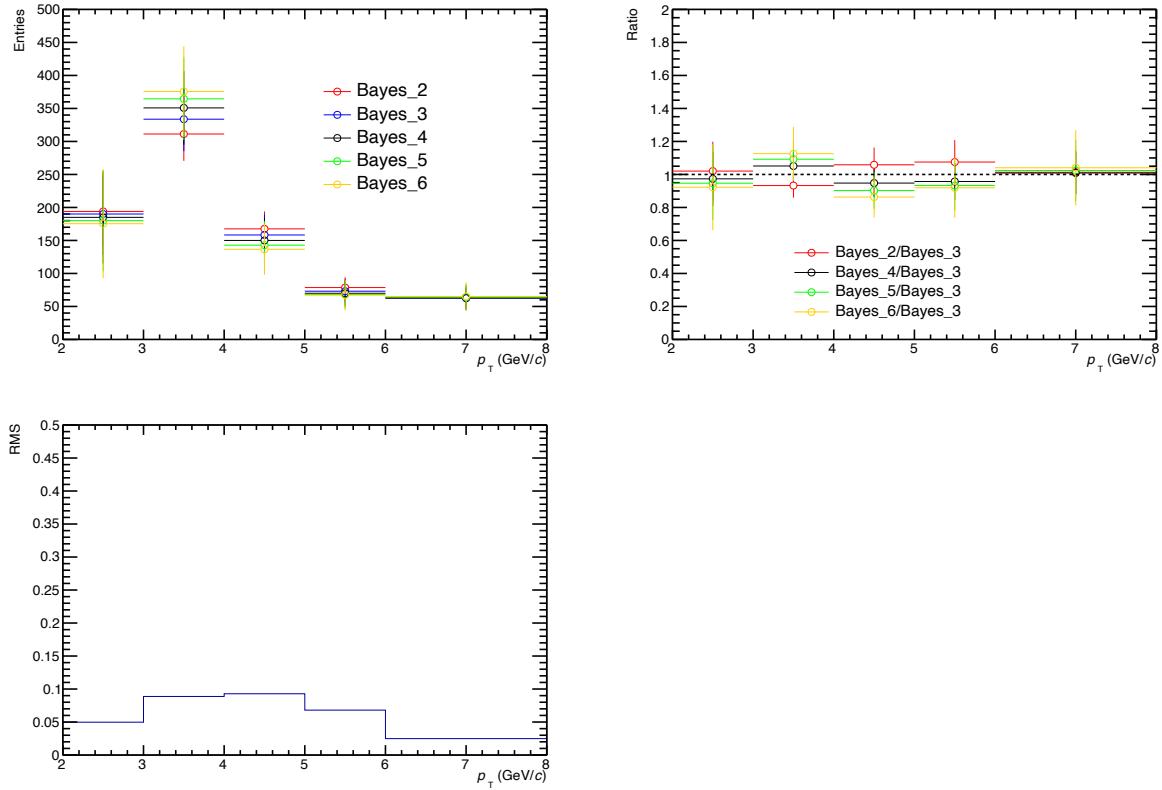


Figure 3.20: The systematic study for the unfolding procedure of varied Bayesian iteration numbers. Top left panel: the p_T distribution of unfolded yield Ξ_c^0 with different Bayesian iteration numbers. Top right panel: ratios of the unfolded yield Ξ_c^0 with variation. Bottom panel: the root mean square (RMS) of the deviations from 1.



3.5.5.2 Singular value decomposition (SVD) method

The Singular Value Decomposition (SVD) method [159] is used to study the effect from different unfolding procedures. The SVD unfolding method is sensitive to the d vector, which determines the regularisation number. To avoid fluctuations, the regularisation numbers 3 and 4 are studied.

The comparison of p_T distribution of unfolded yield Ξ_c^0 from the SVD method compared to the default Bayesian one is shown in the top left panel of Fig. 3.21. A p_T -dependent systematic uncertainty is assigned by the RMS value.

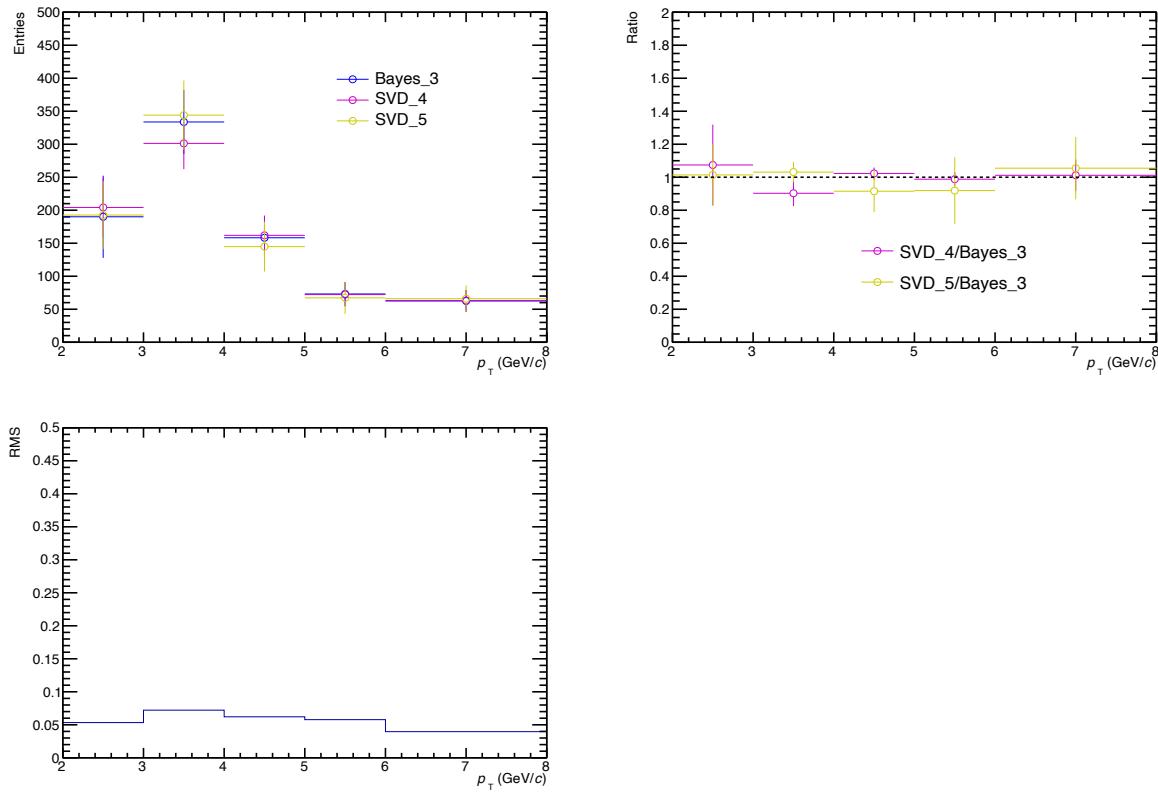


Figure 3.21: The systematic study for the unfolding procedure of Singular Value Decomposition (SVD). Top left panel: the comparison of p_T distribution of unfolded yield Ξ_c^0 from the SVD method compared to the default Bayesian one. Top right panel: ratios of the unfolded yield Ξ_c^0 with variation. Bottom panel: the root mean square (RMS) of the deviations from 1.

3.5.5.3 Unfolding p_T range and binning

The unfolding procedure could be sensitive to the p_T range and the binning, especially at low p_T . To see how much the edge effect, different sets of p_T range and binning of the response matrix are investigated, seen in Fig. 3.22.

Systematic uncertainties of 6% and 4% are assigned in the intervals $2 < p_T < 3$ GeV/c and $3 < p_T < 4$

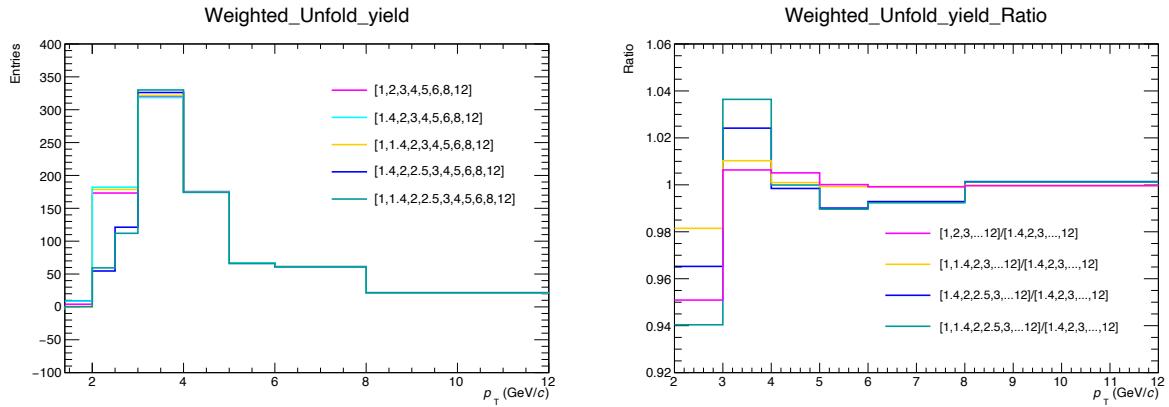


Figure 3.22: The systematic study for the unfolding procedure of p_T binning. Left panel: the p_T distribution of unfolded yield Ξ_c^0 with different p_T binning. Right panel: ratios of the unfolded yield Ξ_c^0 with variation.

GeV/c, respectively. At higher p_T , this contribution is negligible.

3.5.6 Oversubtraction caused by bottom baryons

The assumptions used to subtract the contribution to the WS pairs from $\Xi_b^{0,-}$, already discussed in Section 3.3.2, could bring some uncertainties for the correction. The estimation of systematic uncertainty is considered by varying the $\Xi_b^{0,-}$ yield and momentum distribution based on the uncertainties of the Λ_b^0 p_T -differential cross section in pp collisions [77], listed in the following:

1. Varying the parameters of the Tsallis functions
 - Varying n : varied by 1σ ($7.6 \rightarrow 8.0$)
 - Scale up: CMS estimated the uncertainties on the cross section in the p_T range above 10 GeV/c to be $\approx 50\%$. (The normalization of the Λ_b spectrum is scaled up by the uncertainty)
2. Varying branching ratio fraction
 - The uncertainties on the $B(b \rightarrow \Xi_b B)(\Xi_b \rightarrow e\Xi X)$ (30%) and $B(b \rightarrow \Lambda_b B)(\Lambda_b \rightarrow J/\psi \Lambda)$ (14%) are added in quadrature. The total uncertainty of the branching ratio fraction is 33%.

As shown in Fig. 3.23 and Fig. 3.24, the assigned systematic uncertainty is 1%, independent of Ξ_c^0 p_T .

3.5.7 Prompt fraction estimation

The systematic uncertainty due to the subtraction of the feed-down from beauty-hadron decays, is estimated from two aspects: i) considering the uncertainty on the FONLL predictions; ii) varying the assumption on the ratio Ξ_c^0/Λ_c^+ in the f_{prompt} calculation. Details are listed as follows:

- The uncertainty from FONLL simulations is calculated by varying the b-quark mass and the fac-

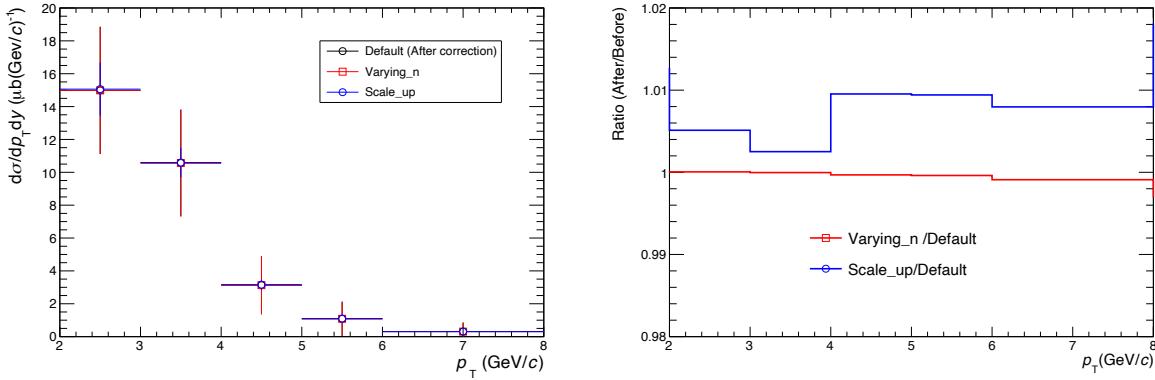


Figure 3.23: The systematic study for oversubtraction by varying the parameters of the TSallis functions.

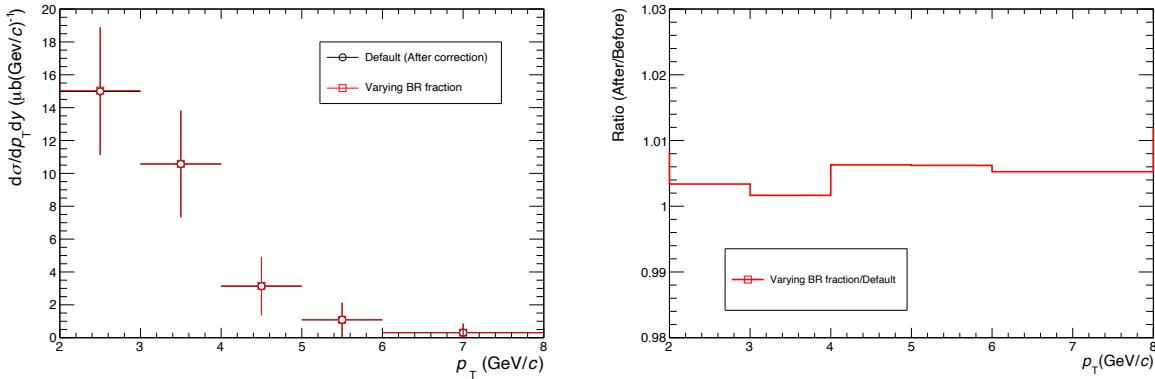


Figure 3.24: The systematic study for oversubtraction by varying branching ratio fraction.

torisation and renormalisation scales as prescribed in Ref. [155], which is shown in the left panel of Fig. 3.25.

- The ratio of inclusive Ξ_c^0 over prompt Λ_c^+ yield, used to multiply the feed-down Ξ_c^0 cross section, is scaled up by a factor of 2 to account for possible differences between the Ξ_c^0/Λ_c^+ and $\Xi_b^{0,-}/\Lambda_b^0$ ratios, and is also scaled down to cover the $\Xi_b^{0,-}/\Lambda_b^0$ value of about 0.12 measured at forward rapidity by the LHCb Collaboration [160], which can be seen in the left panel of Fig. 3.26.

Individual systematic uncertainty from these two sources can be found in Tab. 3.10. The final systematic uncertainty of the subtraction from $\Xi_b^{0,-}$ ranges between 2% and 5%, depending on the p_T interval.

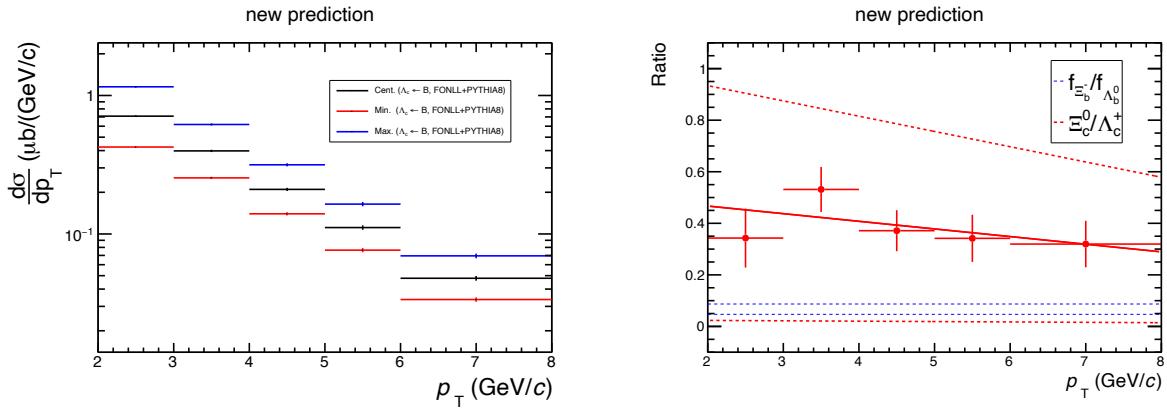


Figure 3.25: The systematic study for prompt fraction estimation.

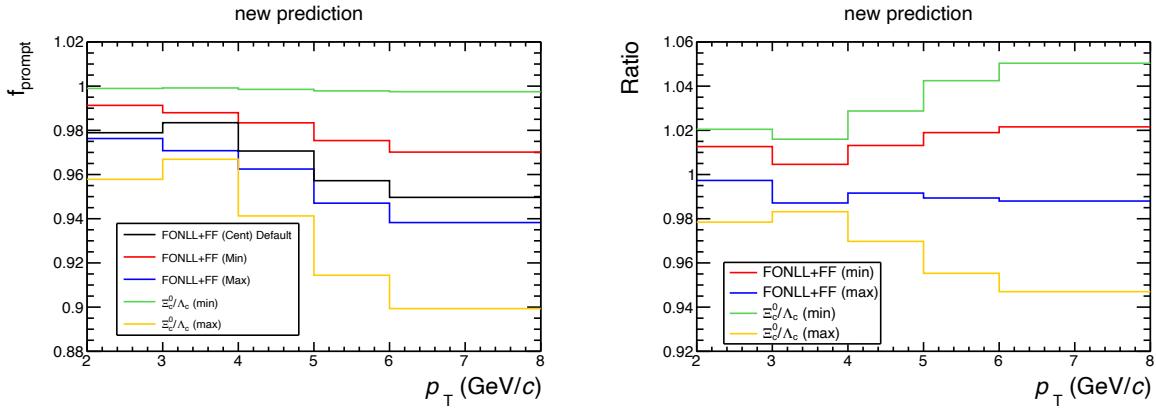


Figure 3.26: The systematic study for prompt fraction estimation. Left panel: The comparisons of prompt fractions obtained from different assumptions. Right panel: The ratio of prompt fractions between the varied ones over the standard one.

Table 3.10: Contributions to the systematic uncertainty of the prompt fraction estimation.

$p_T (\Xi_c^0)$ (GeV/c)		2–3	3–4	4–5	5–6	6–8
FONLL + FF	up	1%	1%	1%	2%	2%
FONLL + FF	down	1%	1%	1%	1%	1%
Ξ_c^0/Λ_c^+	up	2%	2%	3%	4%	5%
Ξ_c^0/Λ_c^+	down	2%	2%	3%	5%	5%
Total	up	2.2%	2.2%	3.2%	4.5%	5.4%
Total	down	2.2%	2.2%	3.2%	5.1%	5.1%



3.5.8 Generated p_T shape

The systematic uncertainty due to the uncertainty of the generated Ξ_c^0 p_T shape used to calculate the efficiency, is studied by using the shape from the PYTHIA 8 generator instead of the one from the iterative procedure. In Fig. 3.27, the left panel shows the comparison of the reconstruction efficiency with (in red) and without (in black) the weight applied (explained above in Fig. 3.12 in Section 3.3.3), and the right panel tells the ratio fitted by a Pol0 function, 2% is assigned as systematic uncertainty.

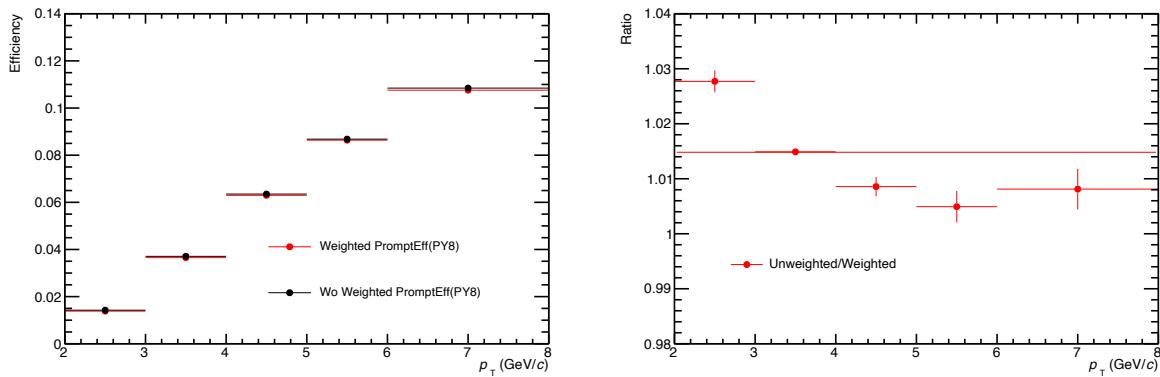


Figure 3.27: The systematic study for the generated p_T shape.

3.5.9 Rapidity range

An additional source of uncertainty is from possible discrepancies with using different rapidity, by comparing the Ξ_c^0 -rapidity distributions in data and in the simulation, which can affect the measured cross section because the $(\text{Acc} \times \varepsilon)$ depends on the Ξ_c^0 rapidity.

The systematic uncertainty is estimated to be 4% by comparing the final corrected yield with different rapidity intervals (from $|y| < 0.5$ to $|y| < 0.8$), shown in Fig. 3.28.

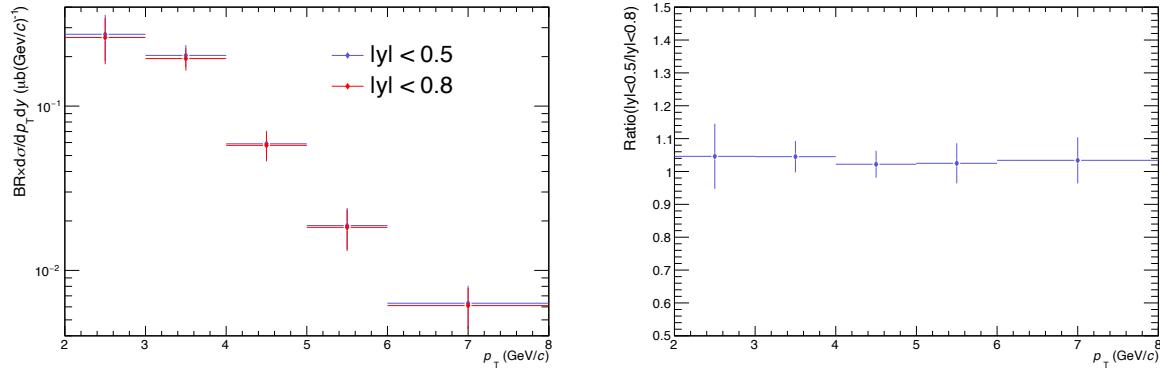


Figure 3.28: The systematic study for rapidity range selection.



3.5.10 Summary of systematic uncertainties

The systematic uncertainty on the Ξ_c^0 production cross section is summarised in Tab. 3.11. The total systematic uncertainty is calculated by summing in quadrature the different contributions, which are assumed to be uncorrelated among each other. Two additional global normalisation uncertainties of the integrated luminosity and the branching ratio are not included in the sum in quadrature with other sources of uncertainties.

Table 3.11: Contributions to the systematic uncertainty of the Ξ_c^0 cross section in this analysis.

p_T (GeV/c)	2–3	3–4	4–5	5–6	6–8
ITS-TPC matching	2%	2%	2%	2%	2%
Electron track selection	2%	2%	2%	2%	2%
Ξ^\pm -daughter track selection	4%	4%	4%	4%	4%
Electron identification	4%	6%	7%	5%	5%
Ξ^\pm topological selection	6%	6%	6%	6%	6%
e Ξ -pair selection	3%	3%	3%	3%	3%
Bayesian-unfolding iterations	5%	9%	9%	7%	2%
Unfolding method	5%	6%	5%	5%	4%
Response-matrix p_T range and binning	6%	4%	—	—	—
Ξ_b oversubtraction	1%	1%	1%	1%	1%
Generated p_T shape	2%	2%	2%	2%	2%
Sensitivity to rapidity interval	4%	4%	4%	4%	4%
Feed-down subtraction	$+2\%$ -2%	$+2\%$ -2%	$+3\%$ -3%	$+5\%$ -5%	$+5\%$ -5%
Total systematic uncertainty	$+14\%$ -14%	$+16\%$ -16%	$+16\%$ -16%	$+15\%$ -15%	$+13\%$ -13%
Branching ratio	28.4%				
Luminosity	2.1%				

3.6 Results

The production of prompt Ξ_c^0 -baryon in pp collisions at $\sqrt{s} = 5.02$ TeV, measured in p_T range $2 < p_T < 8$ GeV/c, is shown in the left panel of Fig. 3.29. It is compared with the previously published results of the inclusive Ξ_c^0 baryon production via its decay mode $\Xi_c^0 \rightarrow \Xi^- e^+ \nu_e$ in pp collisions at $\sqrt{s} = 7$ TeV [34], which is updated with the BR value recently measured from Ref. [21], and of prompt Ξ_c^0 -baryon production at $\sqrt{s} = 13$ TeV [21]. The latter one is the average of two decay channels ($\Xi_c^0 \rightarrow$

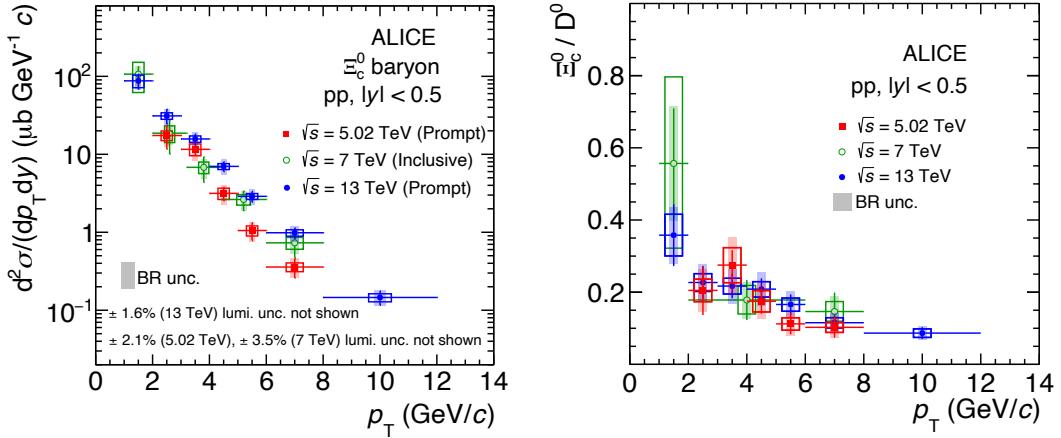


Figure 3.29: Left panel: p_T -differential production cross sections of prompt Ξ_c^0 baryons in pp collisions at $\sqrt{s} = 5.02 \text{ TeV}$ and 13 TeV [21] and of inclusive Ξ_c^0 baryons in pp collisions at $\sqrt{s} = 7 \text{ TeV}$ [34] with updated decay BR as discussed in the text. The uncertainty of the BR of the cross sections of prompt Ξ_c^0 baryons in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ is lower because it consists in the combination of two different decay channels ($\Xi_c^0 \rightarrow e^+ \Xi^- \nu_e$ and $\Xi_c^0 \rightarrow \pi^+ \Xi^-$) [21]. Right panel: Ξ_c^0/D^0 ratio measured in pp collisions at $\sqrt{s} = 5.02 \text{ TeV}$, compared with the measurements at $\sqrt{s} = 7 \text{ TeV}$ [34] and $\sqrt{s} = 13 \text{ TeV}$ [21]. The uncertainty of the BR of D^0 and Ξ_c^0 are shown as shaded boxes. Figures from [33].

$\Xi^- e^+ \nu_e$ and $\Xi_c^0 \rightarrow \Xi^- \pi^+$). The vertical bars indicate the statistical uncertainties and the empty boxes represent the systematic uncertainties. The systematic uncertainties of the BR are indicated as shaded boxes. The data points are positioned at the centre of each p_T interval.

In Fig. 3.14, the prompt fraction in the Ξ_c^0 -baryon yield is close to unity. The comparison of the Ξ_c^0 cross section across the three different collision energies provides an important insight into the \sqrt{s} dependence of the production. It tells that a smaller Ξ_c^0 production cross section is measured at lower collision energies, as expected. The difference among the three results indicates a hardening of the p_T -differential spectrum with increasing collision energy. The same behavior is also observed for measurements of the D-meson and Λ_c^+ -baryon cross sections at $\sqrt{s} = 5.02, 7$ and 13 TeV [8, 10, 20, 96, 116], and with the expectations from pQCD calculations [3, 155].

The p_T -integrated cross section is obtained by integrating the p_T -differential cross section of the measurements, as reported in Eq. 3.14.

$$\frac{d\sigma_{\text{pp}, 5.02 \text{ TeV}}^{\Xi_c^0}}{dy} \Big|_{|y| < 0.5}^{(2 < p_T < 8 \text{ GeV}/c)} = 33.9 \pm 6.0 \text{ (stat.)} \pm 10.6 \text{ (syst.)} \pm 0.7 \text{ (lumi.) } \mu\text{b}, \quad (3.14)$$

In the right panel of Fig. 3.29, the Ξ_c^0/D^0 yield ratio measured in pp collisions at $\sqrt{s} = 5.02 \text{ TeV}$, is compared with the measurements at $\sqrt{s} = 7 \text{ TeV}$ [34] and $\sqrt{s} = 13 \text{ TeV}$ [21]. The systematic uncertainty



on the Ξ_c^0/D^0 yield ratio is obtained with the assumption that all the uncertainties of Ξ_c^0 and D^0 are uncorrelated, except for the tracking and feed-down systematic uncertainties, since they will partially cancel in the ratio. In addition, the uncertainty of the luminosity fully cancels in the ratio. The three Ξ_c^0/D^0 yield ratios are consistent with each other within uncertainties, and they have similar p_T trends. The ratio is close to be 0.2 at low p_T while decreasing with increasing p_T , approximately a value about 0.1 at $p_T > 6 \text{ GeV}/c$. The integrated value of Ξ_c^0/D^0 yield ratio in this analysis, in $2 < p_T < 8 \text{ GeV}/c$, is $0.21 \pm 0.04(\text{stat.}) \pm 0.07(\text{syst.})$, which is obtained by calculating as the ratio of the integrated cross section Ξ_c^0 and D^0 in the same considered p_T intervals.

3.6.1 Comparison with model calculations

The left panel of Fig. 3.30 shows the comparison of the production of Ξ_c^0 baryons with different model predictions. The different tunes of the PYTHIA 8.243 generator, like the Monash tune [35] based on the traditional fragmentation fractions parameterised in e^+e^- collisions, and also three tunes with the implementation of CR beyond leading-colour approximation [23] (discussed above in Section 1.3.2.1). In the PYTHIA 8 simulations, all soft QCD processes are enabled. The Monash tune significantly underpredicts the data by a factor of about 23 at the low p_T interval and by a factor of 5 at the highest p_T interval. This gives additional information about the non-universality of charm fragmentation, as reported in Refs. [11, 20, 21], and also about the consideration that event generators tuned from e^+e^- result can not describe the enhanced baryon production in pp collisions at LHC energies. The three tuned CR models give similar predictions of Ξ_c^0 p_T -differential cross sections, and provide significantly large values compared with the one from the Monash tune. However, they all underestimate the measured Ξ_c^0 production cross section by a factor 5-6 at $2 < p_T < 3 \text{ GeV}/c$, and a factor of around 3-4 at $p_T > 6 \text{ GeV}/c$. The data is also compared with a model using a coalescence approach in hadronic collisions in the framework of QCM [36, 161] (explained in Section 1.3.2.2). The QCM model is closer to the data as compared to those PYTHIA 8 with CR tunes, however, it underpredicts the experimental result by a factor of 2-3 for $p_T < 4 \text{ GeV}/c$.

In the right panel of Fig. 3.30, the measured Ξ_c^0/D^0 ratio is also compared with the different tunes of PYTHIA 8 as explained previously. All PYTHIA 8 tunes give lower predictions for the measured Ξ_c^0/D^0 yield ratio. The Monash tune underestimates significantly the data about a factor 21-24 at low p_T , and a factor of 7 in the highest p_T range, which is also observed for the Λ_c^+/D^0 [11]. For the three CR modes, they give a similar magnitude and shape of the Ξ_c^0/D^0 ratio. Similarly as predicting the production of Ξ_c^0 , they still underestimate the measured baryon-to-meson yield ratio, but predict a larger value than the one from the Monash tune. The models with CR tunes fail to describe the Ξ_c^0/D^0 , however they predicts

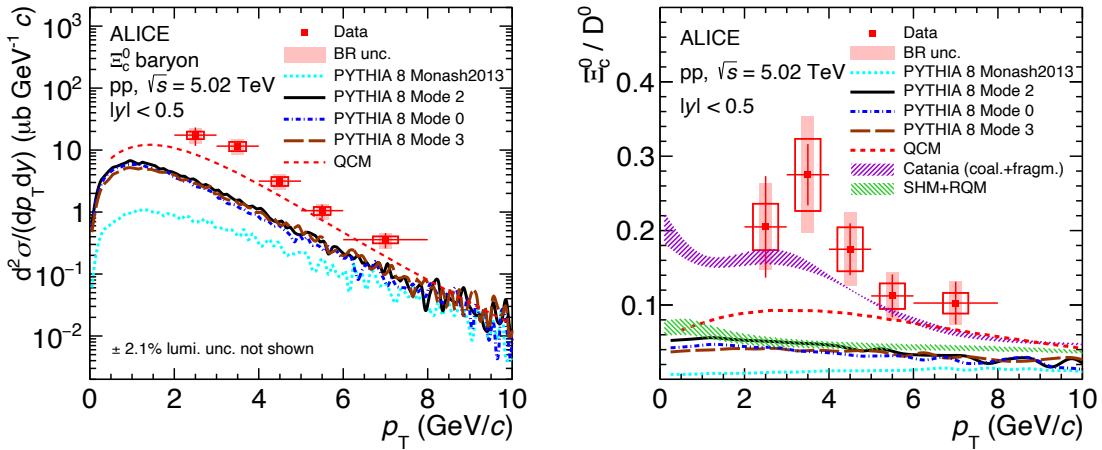


Figure 3.30: Left panel: p_T -differential production cross section of prompt Ξ_c^0 baryons in pp collisions at $\sqrt{s} = 5.02$ TeV compared with model calculations [23, 35, 36]. Right panel: Ξ_c^0/D^0 ratio as a function of p_T measured in pp collisions at $\sqrt{s} = 5.02$ TeV compared with model calculations [23, 24, 35–37] (see text for details).

better for the Λ_c^+/D^0 and the $\Sigma_c^{0,+,\text{++}}/D^0$ ratios [11, 16, 20, 21], as already discussed above in Fig. 1.9 in Section 1.3. One possible reason could be that there is a puzzle about the charm-strange sector.

The data is also compared with the calculation from SHM implemented together with QRM, explained in Section 1.3.2.4. The green uncertainty band shown in the left panel of Fig. 3.30, is obtained by varying the assumption of the branching ratios of excited charm-baryon states decaying to the ground state $\Xi_c^{0,+}$, with the assumption of an exact isospin symmetry between Ξ_c^+ and Ξ_c^0 . It underestimates the Ξ_c^0/D^0 ratio with the same amount as the PYTHIA 8 with CR tunes. However, it is observed to describe for the Λ_c^+/D^0 ratio [11]. The QCM model, explained in Section 1.3.2.2, gives lower predictions for the Ξ_c^0/D^0 ratio by the same amount as it does for the Ξ_c^0 -baryon production cross section. The closest prediction to the data point is from the Catania model, discussed in Section 1.3.2.3. The coalescence process of heavy quarks with light quarks is tuned to have all the charm quarks hadronising via only coalescence at $p_T \approx 0$. The Catania model captures the trend of Ξ_c^0/D^0 ratio in the full p_T interval of the measurement.

The constraint between the Ξ_c^0 measurement and all model predictions mentioned above offers important information on charm quark hadronisation in pp collisions, particularly the role of charm-strange baryon production in the colour reconnection approach.

3.6.2 Extrapolation down to $p_T = 0$ of the Ξ_c^0 cross section and the Ξ_c^0/D^0 ratio

The p_T -integrated production cross section of prompt Ξ_c^0 baryons is obtained by extrapolating the visible cross section to the full p_T range, as reported in Eq. 3.14. The central value of the extrapolation factor is



obtained by the PYTHIA 8 generator with CR Mode 2, adopted what was done for the Λ_c^+ baryon [11]. As the PYTHIA 8 CR Mode 2 captures better the p_T shape of the measured Ξ_c^0 than other models. The p_T -differential Ξ_c^0 cross section values for $0 < p_T < 2 \text{ GeV}/c$ and for $p_T > 8 \text{ GeV}/c$ are obtained by multiplying the measured Ξ_c^0 in this analysis ($2 < p_T < 8 \text{ GeV}/c$) by the ratio of the one predicted with PYTHIA 8 CR Mode 2 in the full and $2 < p_T < 8 \text{ GeV}/c$, respectively. The systematic uncertainty of the extrapolation is estimated from the different extrapolation factors using all available model calculations [23, 24, 36, 37] except for the Monash tune [35], as it can not reproduce the measured Ξ_c^0 cross section. The extrapolation factor is calculated to be $2.65^{+0.54}_{-0.44}$. The p_T -integrated cross section of prompt Ξ_c^0 -baryon production in pp collisions at $\sqrt{s} = 5.02 \text{ TeV}$ is obtained, as Eq. 3.15:

$$\frac{d\sigma_{\text{pp}, 5.02 \text{ TeV}}^{\Xi_c^0}}{dy} \Big|_{|y|<0.5} = 89.8 \pm 16.0 \text{ (stat.)} \pm 28.1 \text{ (syst.)} \pm 1.9 \text{ (lumi.)} {}^{+18.2}_{-15.0} \text{ (extrap.) } \mu\text{b}, \quad (3.15)$$

The result in Eq. 3.15 is used to calculate the ratio to the one of the D^0 meson measured at the same collision energy [16]. The obtained p_T -integrated Ξ_c^0/D^0 ratio is $0.20 \pm 0.04 \text{ (stat.)} {}^{+0.08}_{-0.07} \text{ (syst.)}$. For the systematic uncertainty of the baryon-to-meson ratio, three resources are treated as correlated between the Ξ_c^0 and the D^0 cross sections, the tracking, the FONLL contribution to the feed-down, and the luminosity, and the rest sources are considered as uncorrelated. The extrapolation uncertainty is also included in the total systematic uncertainty. To achieve an accurate measurement of the $c\bar{c}$ production cross section at midrapidity in pp collisions at the LHC, therefore, it is imperative to include the contribution from the Ξ_c^0 baryons.

3.7 Summary

The measurement of the production of prompt Ξ_c^0 baryons in pp collisions at $\sqrt{s} = 5.02 \text{ TeV}$ at midrapidity ($|y| < 0.5$) with the ALICE detector is presented. This analysis is performed using the semileptonic decay channel $\Xi_c^0 \rightarrow e^+ \Xi^- \nu_e$ and its charge conjugate. The p_T -differential cross section is measured in the transverse-momentum interval $2 < p_T < 8 \text{ GeV}/c$.

The measured p_T -differential cross section and Ξ_c^0/D^0 ratio are compared with different model calculations implementing various particle production and hadronisation mechanisms. The PYTHIA 8 simulations with the colour-reconnection mechanism predict an enhanced production of baryons, which is closest to the measured data, compared with the Monash tune.

The p_T -differential Ξ_c^0/D^0 ratio is also compared with several models. The PYTHIA 8 with CR, the



SHM, and the QCM, all fail to capture the measured data. In contrast, the measured Ξ_c^0/D^0 ratio is better described by the Catania model, which implements a new framework allowing low- p_T charm quarks to hadronise also via fragmentation in addition to the coalescence mechanism. The result contributes to compute the total charm cross section at midrapidity in pp collisions at $\sqrt{s} = 5.02$ TeV, as shown in Fig. 1.10, more details can be seen in Ref. [99].

With the enhanced Ξ_c^0/D^0 ratio, this analysis provides additional information on the charm fragmentation, which might not be universal. Furthermore, it offers stringent constraints on models of charm-strange sector hadronisation in pp collisions.



4 Ω_c^0 production in pp collisions at $\sqrt{s} = 13$ TeV

The measurement of the inclusive production of Ω_c^0 baryons, via its semileptonic decay channel $\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e$, is measured for the first time at midrapidity ($|y| < 0.8$) in proton-proton collisions at $\sqrt{s} = 13$ TeV. In this chapter, firstly three new innovative methods implemented in this analysis are explained in Section 4.1. The analysis details, especially about the implementation of ML, as well as the estimation of the systematic uncertainties, are discussed in the next sections subsequently. Section 4.8 presents the results, namely the p_T -differential production cross section of the inclusive Ω_c^0 baryons multiplied by its branching ratio $BR(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e)$, also the branching-fraction ratio $BR(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e)/BR(\Omega_c^0 \rightarrow \Omega^- \pi^+)$, which is compared with theoretical calculations and experimental measurements. Finally, conclusions are drawn in Section 4.9.

4.1 Analysis methods

To improve the precision of these measurements, this analysis is implemented with three new advanced techniques: 1) the Kalman Filter Particle Package [162] used to reconstruct Ω cascade candidates; 2) the machine learning (ML) technique based on the Boosted Decision Trees, to enhance the signal extraction performance; 3) the use of new combinatorial background, so-called mixed-event (ME) technique. The Kalman algorithm has been successfully validated on simulated data of the CBM experiment [163, 164], and also performed in the ALICE experiment [165]. The ML method has gained popularity in high-energy physics for various tasks, including classification and regression problems, and simulation and event reconstruction [166, 167]. The ME technique is used for the signal extraction. In the following, a brief description to these new techniques is presented.

4.1.1 Kalman Filter Particle Package

The traditional method of reconstructing short-lived particles is determined by the secondary vertex of their decay as the point at the DCA between the daughter particle trajectories. Then the mother particle is reconstructed by extrapolating the daughter particles' parameters to that point, in which their momentum and energy are summed up. That is the way how usual vertexing packages reconstruct the production and decay vertices, as employed in the Ξ_c^0 analysis in Chapter 3.

On the other hand, the KFParticle package provides a new method to estimate the decayed particle's parameters and their associated covariance matrix, in addition to the reconstruction of the production and decay vertices [168]. It is a software package developed for the reconstruction of short-lived particles,



with the underlying algorithm based on the Kalman filter method [169].

The Kalman Filter algorithm is a mathematical iterative procedure to estimate unknown variables, taking into account a set of measurements with their inaccuracies. It is intended to find an optimal estimator \mathbf{r} of unknown state vector \mathbf{r}^t with its covariance matrix \mathbf{C} , via taking a set of the measurements m_k , $k = 1 \dots n$ of the vector \mathbf{r}^t . The covariance matrix is a symmetric non-negative definite matrix, whose diagonal elements are the squared dispersions of the corresponding elements of the vector. The algorithm starts with a certain initial approximate value of vector \mathbf{r}_0 and \mathbf{C}_0 , called *initialisation step*. The state vector \mathbf{r}^t will be updated during the iterative steps, with the knowledge of the impact of one measurement on the change of the state vector, a prediction of the values of \mathbf{r}_0 and \mathbf{C}_0 is made based on the first approximation, *prediction step*. The next stage is the essence of the Kalman filter, so-called *filtration step*, for each measurement \mathbf{m}_k an estimator \mathbf{r}_k is calculated via a geometrical fit, which is the optimum estimation of the vector \mathbf{r}_k^t according to the first k measurements. The estimator \mathbf{r}_k^t obtained after the filtration of the last measurement is the desired optimal estimation with the state vector.

The KF method is utilised to estimate the position of a decay vertex and the momentum of each daughter particle at the reconstructed vertex. With the kinematic knowledge, the reconstruction of the mother particle is determined by the position, momentum, and energy of their daughter particles obtained after the track fitting procedure. The KF algorithm uses a geometry-independent particle parameterization as the state vector, expressed as in Eq. 4.2,

$$\mathbf{r} = (x, y, z, p_x, p_y, p_z, E)^T, \quad (4.1)$$

where the position of a particle (x, y, z) , the momentum of a particle $(p_x, p_y, p_z)^T$, and its energy E .

After the first approximation of the decay vertex, the daughter particles are transported to the initial vertex estimation, and then their momentum and covariance matrix are calculated accordingly. Through three-stage estimation performed sequentially, the estimation of the mother particle's state vector and covariance matrix are provided. After the mother particle is reconstructed at the decay point, a new parameter $s = l/p$ can be added to the state vector. It is equal to the particle path length from its production point to the decay point, normalised to the particle momentum, and l is the length of the trajectory in the laboratory coordinate system.

Thus, the state vector is updated as:

$$\mathbf{r} = (x, y, z, p_x, p_y, p_z, E, s)^T, \quad (4.2)$$



With the new parameter s , all parameters are transported from the decay point to the production point, where they are measured by the production vertex. Therefore, the final optimal reconstructed state vector and its covariance matrix contain all necessary information about the particle both at the production point and at the decay point. Additional physics parameters, not explicitly included in the state vector, such as the particle momentum p , the invariant mass M , and the time of life of the particle $c\tau$, can be easily calculated as follows:

$$p = \sqrt{p_x^2 + p_y^2 + p_z^2}, \quad (4.3)$$

$$M = \sqrt{E^2 - p^2}, \quad (4.4)$$

$$c\tau = s \cdot M, \quad (4.5)$$

4.1.1.1 Constrained fits and variables

In some cases, to improve the precision of the state vector estimator obtained from the geometrical vertex fit, it is required several assumptions about the associated tracks [164, 165, 170, 171]. These assumptions are expressed in terms of constraints on the state vector parameters. They are treated as one-dimensional measurements with a null error by the Kalman filter in the secondary vertex fit. Two types of constraints have been included in the vertex fit package: *topological constraint* and *mass constraint* [165, 170].

If an invariant mass M of the reconstructed particle is known, as given in Eq. 4.4, a linearised matrix H^M of the mass measurement can be calculated by taking the partial derivatives of M^2 , followed [172]:

$$H_{M^2} = (0, 0, 0, -p_x, -p_y, -p_z, E, 0), \quad (4.6)$$

The mass constraints are then utilised by the KF as a conventional one-dimensional measurement, with the measured value M^2 , null error, and the measurement matrix H^M :

$$M^2 = E^2 - (p^{x^2} + p^{y^2} + p^{z^2}) + H^M(\mathbf{r}^t - \mathbf{r}^0), \quad (4.7)$$

This constraint penalises the parameters of the state vector parameters, ensuring that all daughter particles are required to form the invariant mass M , so-called *mass constraint*. Consequently, all the mass and momentum information of the reconstructed particle are updated [170]. For analyses involving long decay chains, such as in this analysis, the reconstruction of $\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e$, where the Ω decays into $\Omega^+ \rightarrow K^+ \Lambda$, following $\Lambda \rightarrow \pi^- p$. A mass constraint is set on the Λ just before constructing the Ω^-



will improve a mass resolution of Ω^- , as well as that of Ω_c^0 from Ω^- , where the mass of Λ and Ω^- is constrained to the PDG mass [45].

A constraint can also be set on a particle's production vertex, requiring the particle to originate from that vertex. The topological constraint is used to align a mother particle with its known production vertex. When the particle is required to point back to the production vertex, the uncertainties associated with the vertex can be also taken into account in the topological constraint. In the case that a particle decays close to the production vertex, or in the case of an uncharged mother particle, the trajectory between the vertex and the particle approximates a straight line:

$$v - p \cdot t = v_{pv}, \quad (4.8)$$

with (v, p) position and momentum of the particle, and v_{pv} the position of the production vertex, and an additional parameter t denoting the trajectory length of the mother particle, normalised to its momentum. The particle trajectory is forced to point back to the production vertex. The particle is transported to the production vertex and then gets filtered using the production vertex as a measurement. This constraint can be included directly into the Kalman filter as a set of two independent measurements. Additionally, the algorithm accounts for the errors associated with the production vertex.

A schematic representation of the Ω_c^0 decay, including uncertainties associated with the tracks and vertices of the mother and daughter particles, is depicted in Fig. 4.1.

After the fitting procedure, different quantities related to the decay topology and fit quality can be extracted. These variables are useful to select the specific reconstructed particle candidates. During the reconstruction of particle decay, their daughters are required to follow a geometrical fitting procedure to estimate the estimated production vertex or the decay vertex. The quantity χ_{geo}^2/NDF refers to how well the daughter tracks intersect within their uncertainties, where NDF is the number of degrees of freedom [171]. A small χ_{geo}^2/NDF indicates a high probability of a good vertex fit.

When a reconstructed particle is constrained to point back to the production vertex using a topological constraint, the quantity χ_{topo}^2/NDF is introduced to quantify the probability of the hypothesis that the particle truly comes from the assigned vertex. It characterises whether the particle is produced in the primary vertex region, a smaller obtained value suggests a higher probability that the particle trajectory intersects the primary vertex within the errors. The distance, how far away from the primary vertex a particle decays, is given by the variable $l/\Delta l$. It tells the distance between the decay vertex and the

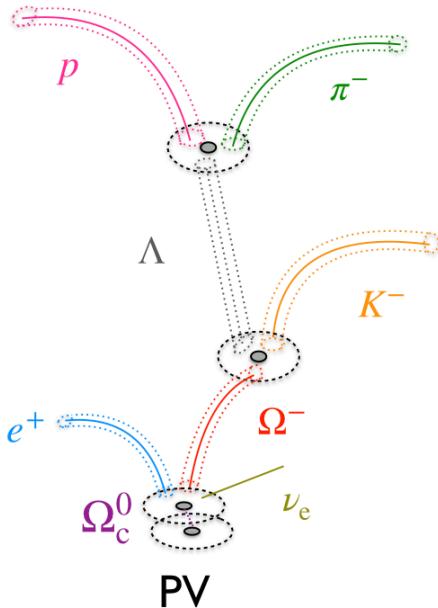


Figure 4.1: Sketch of the semileptonic decay of a Ω_c^0 baryon. The primary vertex (PV) is the collision point where the Ω_c^0 baryon is produced.

primary vertex normalised to the uncertainty on the distance, which are localisation uncertainties of the vertices.

4.1.2 Machine Learning

In the search of rare signals, especially 3-prong decays $\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e$, reconstruction is often challenged by the presence of a huge combinatorial background. The traditional analysis approach, based on “rectangular” selection criteria as in Section 3, may fail a particular selection criterion since some candidates may not have all the characteristics of signal or background. Optimisation of the selection criteria is a delicate task, which is improved by exploiting the popular technique of ML [167, 173] for this analysis. Most common ML methods can be classified into two main categories: *supervised learning* and *unsupervised learning*. In supervised learning, the training data used in the algorithm includes the desired solutions, called *labels*, further explained in Section 4.1.2.1. In the data analysis of High Energy Physics experiments, ML is widely employed to solve difficult classification or regression problems, both belong to supervised learning. In the case that a class label needs to be predicted, called *classification* task, whereas in *regression* problems, a value is to be predicted.

In this analysis, the Boosted Decision Trees (BDTs) algorithm, provided by the XGBoost library [174, 175], is particularly employed for classifying signal and background. The algorithm is trained on labeled classes of signal and background candidates, to distinguish the differences based on their features. It



offers valuable opportunities for physics by taking a more sophisticated approach than standard cut-and-count methods. The left panel in Fig. A.15 demonstrates the advantage of using the ML technique, statistical uncertainties largely improved especially at low p_T , compared to the rectangular selection case.

4.1.2.1 Supervised learning approach for binary classification task

In a classification learning (predictive learning) problem, the dataset D consists of N input *instances* (variables) $x_i = x_1, x_2, \dots, x_n$ with n associated *features*, and a set of class labels $y_i \in 0, 1$, representing the final outcomes to be predicted. A labelled training sample $D = \{x_i, y_i\}_{i=1}^N$ with N instances contains a set of known (x, y) -values. A supervised learning problem assumes there exists a mapping f , relating elements of dataset D ,

$$y_i = f(x_i), \quad (4.9)$$

The learning goal for a classification problem is to approximate f given D using the characteristics of the objects and their classification as the target, resulting in an approximation \hat{f} :

$$\hat{f} = \mathcal{A}(D), \quad (4.10)$$

This approximation can be used to categorise or classify unseen and unlabelled inputs. The algorithm searches for the best approximation of a solution by minimising a specific *loss function* $L(y, \hat{f}(x))$ [176], which measures the quality of the prediction for input variable x_i , based on some metric. The only way to know how well a model will generalise to new samples is to try it out on new cases. To address this, the initial dataset is typically divided into two sets: the *training set* used to train the model, and the *test set* to check the performance of the model.

However, the ML algorithms are subject to *overfitting* and *underfitting problems*, which both lead to poor model performance. By far the most common problem applied in ML is overfitting. It occurs when the model performs well on the training data, but fails to predict new observations in another dataset. This is illustrated in Fig. 4.2, where the red points are the background and the blue represents the signal. The black curve depicts the separation between the signal and the background. It captures the characteristics of the data set, but it does not fit every fluctuation in the signal. This is different from the behavior of the green line, which catches even fluctuations in the signal. To mitigate overfitting influence and evaluate ML algorithms effectively, many different ways can be employed, as discussed later in Section 4.3.3.

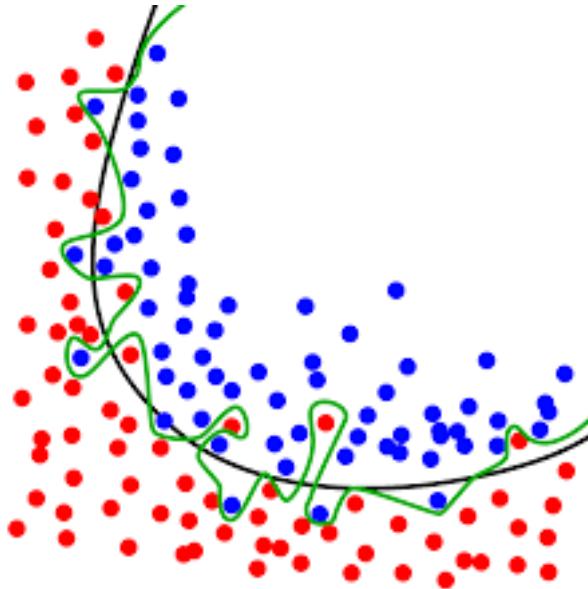


Figure 4.2: Example of the separation between signal (blue) and background (red). the black line fits the data well, the green line is overfit.

4.1.2.2 Boosted decision trees and XGBoost

Boosted Decision Trees (BDTs), are one of the ML robust classification tools [177], and are particularly applicable to missing values, imbalanced datasets, and redundant attributes, concerning their advanced ability [178, 179]. These advanced abilities to deal with tricky cases make them a favored tool in high-energy physics.

A basic structure of a single decision tree is shown in Fig. 4.3. The procedure begins with a set of input features, *initial node* containing all labeled candidates, one of the nodes is then split into two parts, called *branches*, depending on the value of these features. The further splitting procedure is iteratively applied to each branch, to maximise the separation between the signal and background. To prevent overfitting, *pruning* [180] technique may be employed when the gain of a branch smaller than a certain value is not added. As the tree grows, each node contains fewer events, resulting in increased statistical uncertainty with each new split. Ultimately, a given number of final branches called *leaves*, also called *terminal node*, represents class labels or probabilities. Signal-dominated branches are classified as a signal, while background-dominated branches are classified as background [177]. This simplified description outlines the iterative processes of how a decision tree is built and highlights its effectiveness in classification tasks.

Since the 1980s [181], simple decision tree-based algorithms have been successfully applied in data

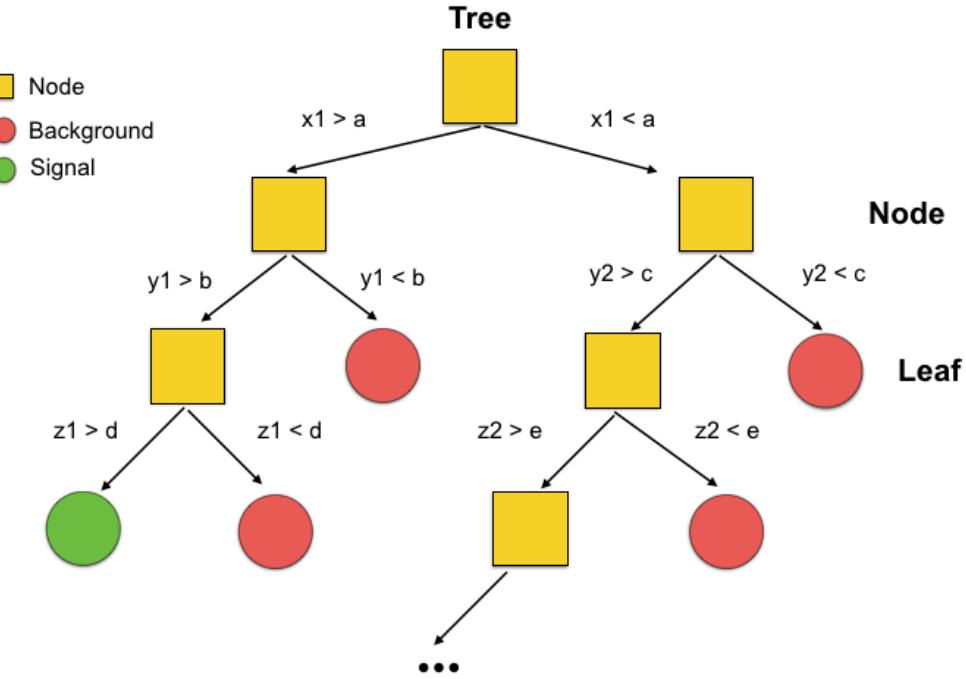


Figure 4.3: Example of a decision tree structure for binary classification of signal (S , in green) and background (B , in red) candidates. The initial sample at the first node is split into branches. The terminal nodes are shown as circles and are either signal leaves in the case that the signal candidates are dominant, and vice versa for the background leaves.

analysis. However, due to their unstable features, known as *weaker learners*, these algorithms may not perform well as other standalone classifiers(algorithms). To address this limitation, a new algorithm called *boosting* [177] is introduced. Boosting is a procedure that combines many weak classifiers to achieve a powerful multivariate classifier, which can be applied to any classification method. In this analysis, it is specifically applied to decision trees.

In boosting, the process begins with a sample of unweighted candidates, as described in Fig. 4.3. At each step of node splitting, candidates that are misclassified, such as signal candidates ending up in the background leaf, or vice versa, are identified and their weights are increased, hence the term *boosted* [178] is introduced. Subsequently, a new decision tree is built based on these updated weights. This iterative process continues, with each iteration k improving the classification of samples that are misclassified in previous steps. The strength of boosting depends on the ensemble of all the weak learners collectively, contributing to the final classification. In this binary classification case, the outputs (scores) are typically assigned values of 1 for signal and 0 for background.

A generic boosting algorithm [182] exploits many of *weaker learners* C_k into a new, more stable learner with a reduced error rate and improved performance. Each new learner corrects the mistake from the

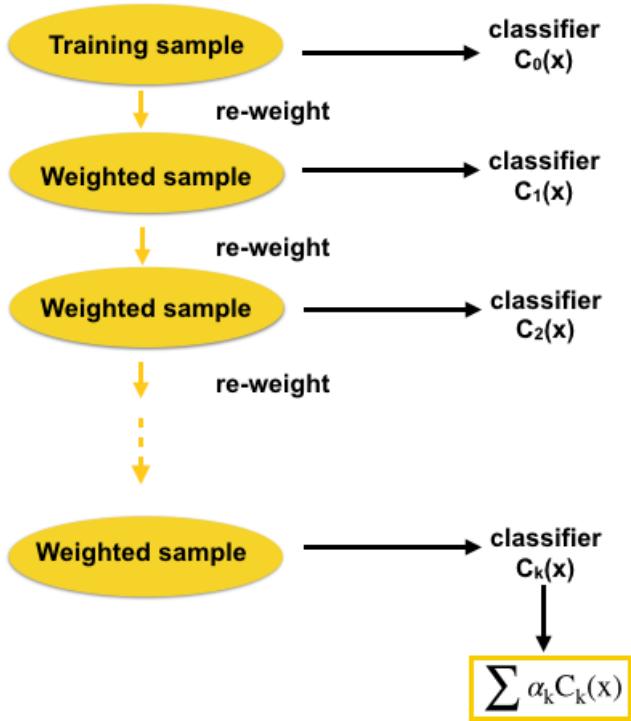


Figure 4.4: Schematic of a boosting procedure.

previous one, leading to a refined classification model. The output of the boosting algorithm is typically a weighted average.

$$\hat{y}_i = \sum_{k=1}^{N_{\text{tree}}} \alpha_k C_k(x_i), \quad (4.11)$$

where N_{tree} represents the number of decision trees built, α_k denotes the weights assigned to the i th event, and $C_k(x_i)$ signifies the prediction made by the k -th tree for the i -th candidate, based on its input features x . Each candidate undergoes scoring by all trees, and the overall score is obtained as the renormalised sum of all the scores, potentially including weights [177]. A high score denotes a large probability of the candidates being signals, or vice versa.

The optimization of the ensemble of weak learners involves minimising the *regularized objective function* \mathcal{R} , as shown in Eq. 4.12. This function takes into account both the *training loss* L and a *regularisation term* Ω , to balance between predictive performance and model complexity.

$$\mathcal{R}^k = \sum_{i=1}^N L(y_i, \hat{y}_i + \alpha_k C_k(x_i)) + \Omega(C_k), \quad (4.12)$$



The training loss function L evaluates the model's predictive accuracy on the training data, while the regularization term controls the complexity of the model, which helps to avoid overfitting.

The gradient descent algorithm [183], introduced to minimise the training loss function, iteratively adjusts parameters in the direction of the negative gradient, to find the optimal set of parameters. Gradient boosting is commonly used in decision trees of fixed size as base learners. XGBoost [180], which stands for “Extreme Gradient Boosting”, is an extension of generic gradient boosting methods, and is employed for binary classification tasks. It leverages a decision tree ensemble model comprising several classification trees.

Before running XGBoost, parameters must be configured. There are different types of parameters: general parameters, which dictate the choice of a booster, in this case, models are tree-based; and booster-specific parameters, which are specific to the selected booster. Here are some key booster-specific parameters:

- **Maximum depth:** Determines the number of nodes in the tree. Increasing the value will make the model more complex and more likely happen to overfit.
- **Learning rate:** Controls the magnitude of change between sequential trees. It quantifies how much the weights are adjusted and controls the step size of the gradient boosting. Smaller values require more iterations but can improve generalisation. A good choice of learning rate can significantly reduce the number of iterations needed until an approximate solution is found.
- **Number of estimators:** Specifies the number of decision trees used in the ensemble model.
- **Minimum child weight:** Sets the minimum sum of weight required to create a new node in the tree. If the tree partition step results in a node where the sum of the instance weights is less than the value of this parameter, then, the building process will give up further partitioning.
- **subsample:** Defines the fraction of training samples used to train each tree.
- **colsample-bytree:** Specifies the fraction of features used to train each tree.

In the heavy-ion physics environment for machine learning, (hipe4ml) [175], booster parameters can be optimised using an *optuna* approach, discussed later in Section 4.3.2. Once parameters are set and the dataset is prepared, the model can be trained and applied. After training, the model can provide predictions on the dataset, in the form of scores or output probabilities, as explained above. Proper application of the model is essential to separate the signal from the background, more discussions are explained in Section 4.4.



4.1.3 Combinatorial background techniques

To increase statistics in the background sample, two new methods are exploited, which are mixed-event (ME) and track rotation (TR). At the beginning of the performance study, rectangular selections are used, which are listed in Appendix A.2.1.1. The raw yield is obtained by subtracting the combinatorial background from the right sign (RS) in the same event (SE), as illustrated in Section 3. These approaches give direct comparisons among the different background subtraction techniques. In the following, ideas and results about the two new background methods are presented.

4.1.3.1 Mixed-Event method

The idea of ME is to create pairs of tracks to mimic uncorrelated backgrounds. The mixed-event correlation distribution is obtained by correlating Ω candidates in an event with electron candidate tracks from other events with similar multiplicity and primary-vertex position along the beam direction. Events are grouped into *pools* based on the primary vertex position along z and the event multiplicity (i.e. the number of tracklets in the SPD with $|\eta| < 0.8$). Once a pool is filled with 200 events, the electron tracks from a given event are mixed with Ω candidates from other events. The configuration of the pool for mixing is defined as follows, Eq. 4.13:

$$\begin{aligned} \text{Mult.}[] &= \{0, 5, 10, 15, 20, 25, 30, 40\}, \\ \text{Vtxz}[] &= \{-10, -7.5, -5, -2.5, 0, 2.5, 5, 7.5, 10\}, \end{aligned} \tag{4.13}$$

The SEWS $e\Omega$ pairs are used to properly normalise the ME $e\Omega$ pairs background sample. A comparison of the invariant-mass $e\Omega$ pairs is shown in Fig. 4.5. It tells that the distribution from ME matches well with the WS pairs in the same event.

4.1.3.2 Track rotation method

The second approach to studying background involves creating pairs of tracks within the same event and rotating one of the tracks by a fixed angle to break possible correlations. Subsequently, the invariant mass of these pairs is recalculated. In this analysis, electron tracks are rotated by a fixed angle, and new pairs are formed within the same events. The configuration of track rotation includes 19 steps of $\pi/10$, ranging from $\pi/10$ to $19\pi/10$.

Fig. 4.6 shows the comparison of invariant mass distribution of $e\Omega$ pairs between SE and TR in three different p_T intervals.

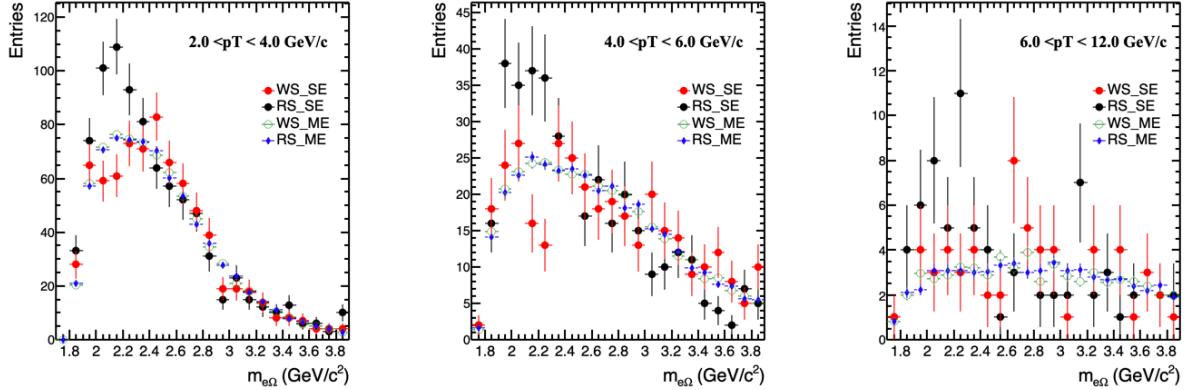


Figure 4.5: Comparison of invariant mass distribution of $e\Omega$ pairs between SE and ME in three p_T intervals.

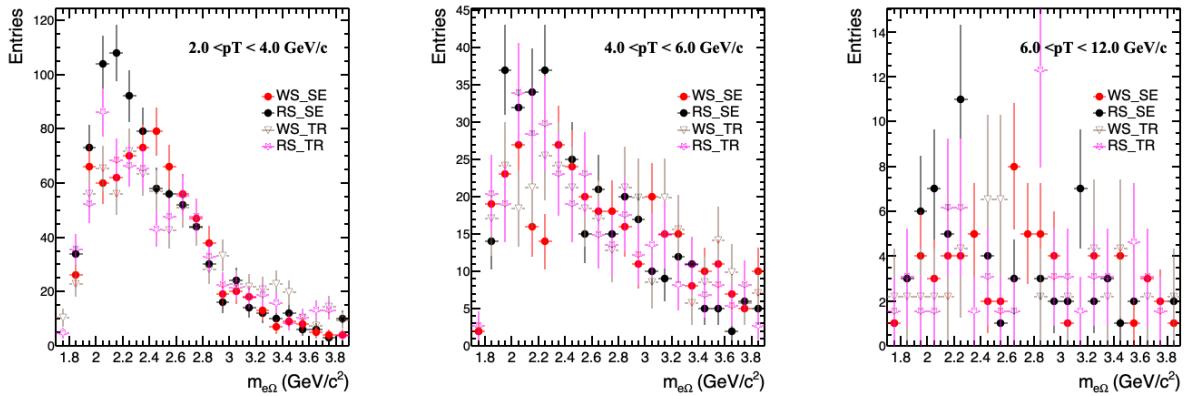


Figure 4.6: Comparison of invariant mass distribution of $e\Omega$ pairs between SE and TR in three p_T intervals.

Additionally, Fig. 4.7 provides an overall comparison of the invariant-mass distribution of $e\Omega$ pairs from SE, ME and TR. The distribution from track rotation, mixed events, and wrong-sign pairs appear compatible within uncertainties. It clearly shows that the mixed-event method yields higher statistics compared to track rotation, regarding the statistical uncertainty. Therefore, the mixed event is chosen for the next step.

4.2 Candidate selection and reconstruction

The main detectors used in this analysis are the Inner Tracking System (ITS), the Time Projection Chamber (TPC), and the Time-Of-Flight detector(TOF) for vertexing, tracking, and particle identification (PID) purposes. They are located in the central barrel, covering the pseudorapidity interval ($|\eta| < 0.9$), and lie inside a solenoidal magnet that provides a magnetic field $B = 0.5$ T parallel to the beam direction. The analysed data sample consists of pp collisions at $\sqrt{s} = 13$ TeV, recorded with a minimum-bias (MB)

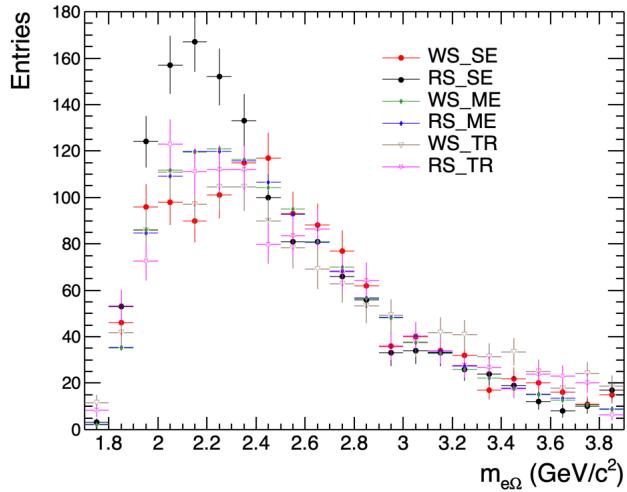


Figure 4.7: Comparison of invariant mass distribution of $e\Omega$ pairs in the interval $2 < p_T^{e\Omega} < 12 \text{ GeV}/c$.

trigger by ALICE collected in the years 2016, 2017, and 2018 during the Run 2 of the LHC. Further offline event selection was applied to remove the contamination from beam-gas collisions and other machine-related backgrounds. Only events with a reconstructed vertex position within $\pm 10 \text{ cm}$ along the beam axis from the nominal interaction point were analysed, to maintain a uniform ITS acceptance in pseudorapidity. With all these selections, a total number of 1.8 billion events are used in this analysis, corresponding to an integrated luminosity of $\mathcal{L}_{\text{int}} = (32.08 \pm 0.51) \text{ nb}^{-1}$ [184].

The analysis also needs the use of signal generated in Monte Carlo (MC) simulations for several purposes, including the ML training, the response matrix used in the unfolding procedure, and also the determination of the reconstruction efficiency. For the reconstructed signal, $e\Omega$ pairs are obtained from simulations using the PYTHIA 8.243 event generator [13]. Each PYTHIA event is required to contain a $c\bar{c}$ or $b\bar{b}$ quark pair and Ω_c^0 baryons are forced to decay into the $\Omega^- e^+ \nu_e$ channel. The mean proper lifetime of Ω_c^0 baryons in the simulation is set to 268 fs based on the latest LHCb measurement [185]. The transport of simulated particles within the detector is performed with the GEANT 3 package [157]. The conditions of all the ALICE detectors in terms of active channels, gain, noise level, and alignment, as well as the evolution of the detector configurations during the data taking period, are taken into account in the simulations. There are about 49 million MC events used in this analysis.

Ω_c^0 candidates are built by pairing an electron (positron) candidate track with an Ω cascade candidate using a Kalman-Filter (KF) vertexing algorithm [134]. Charge conjugate modes are included everywhere, unless otherwise stated. The Ω candidates are reconstructed via the decay chain $\Omega^- \rightarrow \Lambda K^-$ ($\text{BR} =$



Table 4.1: Tracking cuts for electrons applied in this analysis.

Cuts variables	cuts
AOD Filter Bit	4(Standard cuts with very loose DCA)
Number of CrossedRows	>70
CrossedRows Over Findable Cluster	>0.8
Number of TPC PID clusters	>50
Number of ITS cluster	>3
ITS/TPC refit	TRUE
p_T^e (GeV/c)	>0.5
$ \eta $	< 0.8
SPD hit	kBoth (suppress photon conversion)
prefilter cut	$m_{e^+e^-} < 0.05 \text{ GeV}/c^2$

Table 4.2: The tracking cuts for Ω daughters applied in this analysis.

Cuts variables	cuts
Number of CrossedRows	>70
CrossedRows Over Findable Cluster	>0.8
Number of TPC PID clusters	>50

Table 4.3: The further pre-selections applied in this analysis.

Cuts variables	$2 < p_T (\text{GeV}/c) < 4$	$4 < p_T (\text{GeV}/c) < 6$	$6 < p_T (\text{GeV}/c) < 12$
$n\sigma_{\text{TOF}} (\text{K} \leftarrow \Omega)$	-999 or (-5, 5)	-999 or (-5, 5)	-999 or (-5, 5)
$n\sigma_{\text{TOF}}(\text{e})$	(-5, 5)	(-5, 5)	(-5, 5)
$n\sigma_{\text{TPC}}(\text{e})$	(-4, 4)	(-4, 4)	(-4, 4)
CosOA	>0	>0.25	>0.5
$\chi^2_{\text{topo}} (\Omega \text{ to PV})$	>0	>0	>0

$(67.8 \pm 0.7)\%$, followed by $\Lambda \rightarrow p\pi^-$ ($\text{BR} = (63.9 \pm 0.5)\%$) [45], exploiting the characteristic decay topologies as reported in Refs. [25, 151]. Charged-particle tracks used in this analysis are required to be within the pseudorapidity interval $|\eta| < 0.8$ and to have a number of crossed TPC pad rows larger than 70 out of a maximum of 159. Particle identification (PID) selections are based on the differences between the measured and expected response for a given particle species hypothesis, in units of the detector resolution ($n\sigma^{\text{det}}$), explained in Section 2.2.2 and Section 2.2.3. For proton, pion, and kaon



tracks, a selection on the measured specific energy loss dE/dx in the TPC of $|n\sigma_{\text{TPC}}| < 4$ is applied for the respective particle hypothesis. An additional PID selection of $|n\sigma_{\text{TOF}}| < 5$ is applied for the kaon candidates when a measurement of the TOF detector is available. Tracks without TOF hits are identified using only the TPC information. For the KF variables, described above in Section 4.1.1.1, the Ω candidates are required to have a $\chi^2_{\text{topo}}/\text{NDF} > 0$, to remove candidates in which numerical instabilities are observed in the KF fitting procedure. This variable calculated by the KF vertexing algorithm and characterises whether the Ω^- candidate points back to the primary vertex, being the NDF the number of degrees of freedom involved in the track fit procedure.

The first selection for electron candidate tracks is shown in Tab. 4.1, which is similar to the Tab. 3.1 in Section 3.2.1. They are required to have a minimum of 3 hits, out of a maximum of six, in the ITS among which two in the SPD layers [147, 148], at least 50 clusters in the TPC, a number of crossed TPC pad rows larger than 70, and $p_T > 0.5 \text{ GeV}/c$. These requirements help suppressing the contribution from short tracks, which are unlikely to originate from the primary vertex. The dominant source of electron background is photon conversions. By requiring hits in the innermost SPD layers the effective material budget is minimized because conversions further out are rejected by this selection. Two additional selection criteria on the PID of electron candidates, $|n\sigma_{\text{TPC}}^e| < 4$ and $|n\sigma_{\text{TOF}}^e| < 5$, are required. Electrons from conversion occurring in the beam pipe or inside the SPD material can pass the primary track selection criteria, which accounts for a very small fraction. The remaining electrons coming from photon conversion and those originating from Dalitz decays of neutral mesons are rejected with an invariant-mass technique [149, 150], called *prefilter*. The electron candidates are paired with opposite-sign tracks from the same event passing loose identification criteria ($|n\sigma_{\text{TPC}}^e| < 5$ without any TOF requirement) and are rejected if they form at least one e^+e^- pairs with an invariant mass smaller than $50 \text{ MeV}/c^2$. Electron-candidate tracks without TOF PID information are not included in this analysis, as it will bring huge contamination from other hadrons.

Ω_c^0 candidates are selected by requiring the cosine of the opening angle between the electron and the Ω candidate tracks (CosOA) tracks to be greater than 0 for $2 < p_T < 4 \text{ GeV}/c$, 0.25 for $4 < p_T < 6 \text{ GeV}/c$ and 0.5 for $6 < p_T < 12 \text{ GeV}/c$. The p_T dependence of this selection is chosen by looking at its correlation with the $e\Omega$ -pair mass distribution in data and MC, minimising the rejection of signal candidates in the data. All the pre-selections are listed in Tab. 4.1, 4.2 and 4.3, to suppress the combinatorial background at first stage of this analysis.



4.3 Binary classification with XGBoost

After applying the pre-selections described above, a rough separation of signal and background can be obtained. In this analysis, the classification can be further achieved by exploiting the gradient boosting machine XGBoost, described above in Section 4.1.2.2, resulting in the use of a BDT model. With the ML approach, multiple selection criteria are combined into a single response variable. This is the difference with respect to the traditional method, where rectangular selections are applied for various criteria separately. Independent BDT models are trained, with a sample of signal and background candidates for each p_T interval, as listed in Tab. 4.4. For the training inputs: 1) reconstructed signal, as described above in Section 4.2, and 2) background sample obtained using a mixed-event technique (introduced in Section 4.1.3.1), by exploiting randomised sub-samples of the full dataset, using the same filtering selections (described in Tab. 4.1, 4.2 and 4.3), generating $e\Omega$ pairs with the opposite charge. The proportion of signal to background candidates is set to 1 : 2. A proportion 1 : 1 would be the natural choice if enough training candidates are available for good training performance. However, the number of signal candidates from the MC sample is limited, and there are significantly more background candidates. The exact number of the different types of candidates used for training is listed in Tab. 4.4. The data set is then randomly divided into two parts: 60% for training and 40% for testing models' performance.

4.3.1 Training variables

The BDT training variables are topological properties of the decays and PID variables. When choosing these classification criteria, a few concerns have to be taken into account. The number of training variables has an impact on the performance of XBoost algorithm. With an increasing number of variables, the model gains performance in signal and background separation but also becomes complex. It can lead to overtraining, which becomes visible in a larger difference between the performance of the training and test set. Only those variables having the largest impact on the model performance should be included, to get a better model. The most important point is that those variables having a strong correlation with our signal should be excluded in the training procedure. In this analysis, they are the mass of the $e\Omega$ pairs, p_T of $e\Omega$ pairs, and the CosOA between e and Ω candidate tracks.

Table 4.4: Number of signal and background candidates used for the BDT model training and testing in the analysed p_T intervals.

p_T (GeV/c)	(2, 4)	(4, 6)	(6, 12)
Signal	48614	52936	40413
Background	97228	105872	80826



$$n\sigma_{\text{comb}}^{\text{e,K}} = \begin{cases} |n\sigma_{\text{TPC}}^{\text{e,K}}| & \text{tracks only in TPC,} \\ |n\sigma_{\text{TOF}}^{\text{e,K}}| & \text{tracks only in TOF,} \\ \frac{1}{\sqrt{2}} \sqrt{(n\sigma_{\text{TPC}}^{\text{e,K}})^2 + (n\sigma_{\text{TOF}}^{\text{e,K}})^2} & \text{tracks in TPC and TOF,} \end{cases} \quad (4.14)$$

The training variables related to the PID information are the combined number of σ from the TPC and TOF, $n\sigma_{\text{combined}}^{\text{K,e}}$, in Eq. 4.14, on the electron and Kaon tracks, respectively. The training variables describing the Ω^- decay topology are: i) the distance of the closest approach (DCA) of the charged decay particles; ii) the pointing angle of the reconstructed Ω momentum to the primary vertex; iii) the $\chi^2_{\text{topo}}/\text{NDF}$, which characterises whether the momentum vector of the Ω candidate points back to the primary vertex; iv) the $\chi^2_{\text{geo}}/\text{NDF}$, which is defined as the intersection probability of the daughter-particle trajectories, taking their uncertainties into account. The training variables related to the Λ are: a) the DCA to the primary vertex; b) the radial distance of the Λ decay vertex from the beam axis; and c) the DCA between the decay particles. The BDT model output is a single response variable related to the probability that the candidate is a signal.

The signal and background distribution for p_T interval $2 < p_T < 4 \text{ GeV}/c$ is shown in Fig. 4.8. The distribution comparison for the rest p_T intervals can be found in Appendix A.2.4. Fig. 4.9 depicts the correlation matrix for all selected classification criteria in the signal and the background samples for p_T interval $2 < p_T < 4 \text{ GeV}/c$. The correlation matrices for the rest p_T intervals can be found in Appendix A.2.4. As we expect that there is no correlation between the training variables with the three variables that are not used in the training, which is an important criterion for the selection of training features.

4.3.2 Hyperparameters

A model hyperparameter is a configuration external to the model and cannot be estimated from the data. These hyperparameters are used to control the learning process of the ML model. Apart from grid or random grid searches, which can be extremely hard to describe in these frameworks for large-scale experiments. In this analysis, a next-generation hyperparameter optimisation framework called *Op-tuna* [186, 187], which employs Bayesian optimisation with an algorithm known as *TPE* (Tree-structured Parzen Estimator).

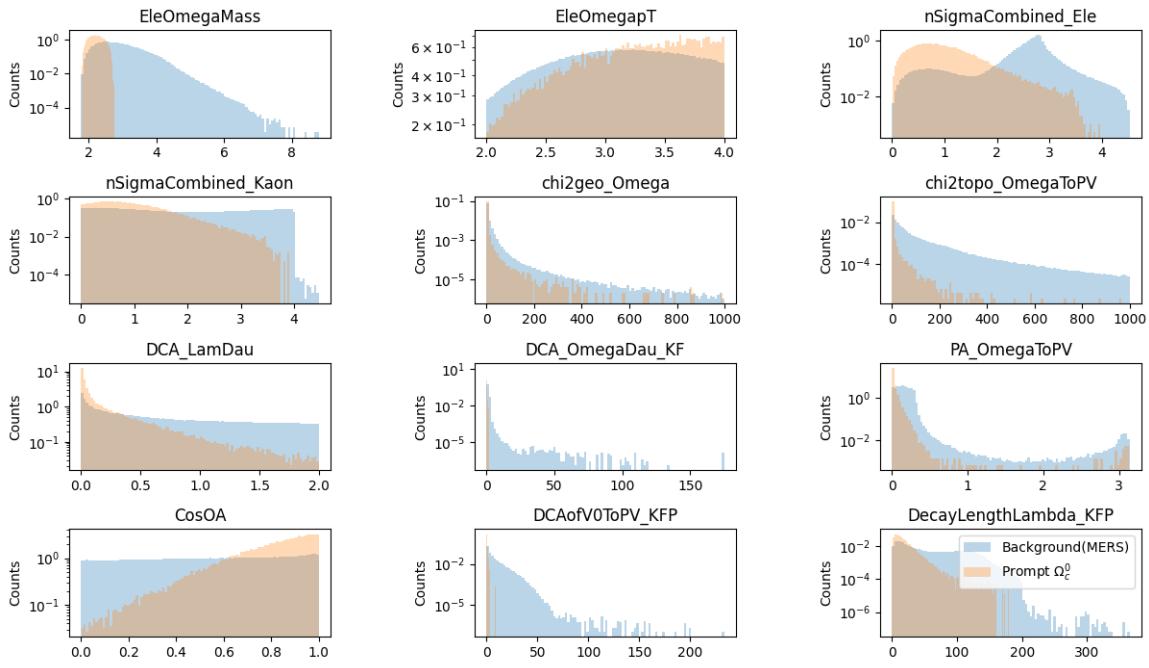


Figure 4.8: Training variables (expect EleOmegaMass, EleOmegapT, CosOA) distribution in $2 < p_T < 4$ for signal (reported in orange) and background (reported in blue) candidates.

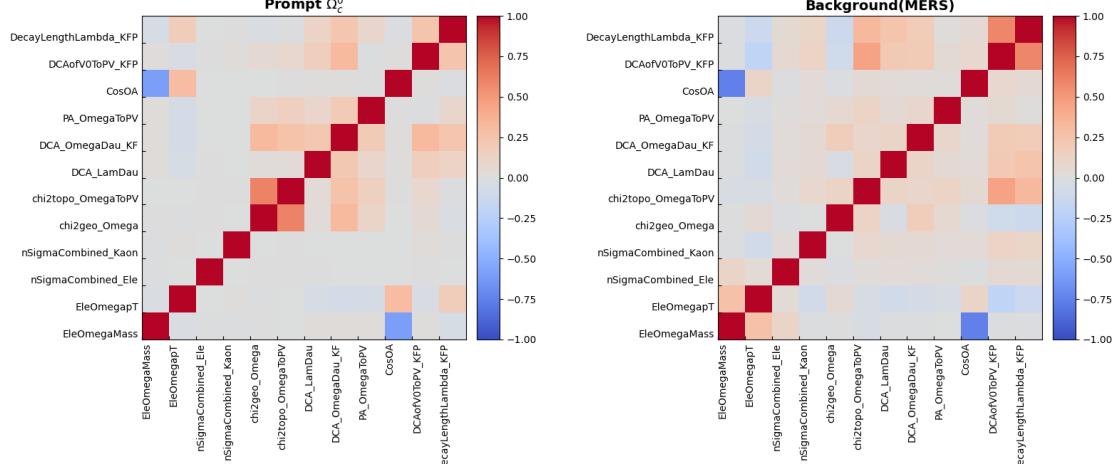


Figure 4.9: Correlation matrix in the transverse momentum range $2 < p_T < 4$.

Optuna is an automatic hyperparameter optimisation software framework, specifically designed for ML. It allows users to dynamically construct the parameter search space using an imperative and unique *define-by-run* API (Application Programming Interface). The idea of *Optuna* is to formulate hyperparameter optimisation as a process of minimising or maximizing an *objective function*, as explained in Section 4.1.2.1, which takes a set of hyperparameters as an input and returns its validation score. In *Optuna*, each optimisation process is referred as a *study*, and each evaluation of the objective function



is called a *trial*. A *study* needs a function to optimise and consists of multiple trials. Each *trial* object corresponds to a single execution of the objective function and is internally instantiated upon each invocation of the function. Therefore, *Optuna* builds the objective function through the interaction with the *trial* object, dynamically constructing the search spaces and generating hyperparameters for each trial accordingly. Unlike random or grid-like searches, the iterative Bayesian Optuna approach considers past evaluations when choosing the next trial. In order to test the hyperparameters, a cross-validation method called *k-fold* [188] (the number of folds to calculate cross-validation error) is used. For this reason, the original sample will be split randomly into k folds and each parameter set on $k-1$ folds will be evaluated. The remaining one is treated as the *validation set*. After testing each hyperparameter set, the folds are permuted for optimisation and validation. The final cross-validation estimation is the mean value of k permutations.

This analysis employs the Bayesian optimisation method with *k*-fold cross-validation implemented in XGBoost library [175, 180]. The hyperparameter space has to be chosen carefully to avoid overtraining and excessive memory consumption. It is important to ensure that the parameter ranges for the optimisation don't always converge to the lower or upper edge of the given range, as this could result in the optimal set being out of the given range. For example, particularly deep trees are resource-intensive, consume significant memory, and introduce possible overfitting, despite improving the performance of the model. Overfitting can be mitigated by setting thresholds for parameters such as the maximum depth of a tree, learning rate, as introduced in Section 4.1.2.2. The optimised parameter sets for each p_T interval used in the BDT models, listed in Tab. 4.5, are applied to extract the final results in this study.

Table 4.5: Optimised hyperparameters used in the analysed for each p_T interval.

p_T (GeV/c)	(1, 2)	(2, 4)	(4, 6)	(6, 12)
Maximum depth	3	3	2	3
Learning rate	0.042	0.039	0.064	0.094
Number of estimators	767	865	846	753
Minimum child weight	3	3	6	3
subsample	0.977	0.826	0.962	0.998
cosamplebytree	0.85	0.977	0.912	0.857



4.3.3 Machine Learning Models

The BDT models are trained for each p_T interval, followed by testing on an independent test set to evaluate their performance. If the performance is optimal, the model will be expected to be neither overtraining nor undertraining, with negligible deviation between the training and test sets. This can be verified with the learning curves, which display the performance of an ML model as the number of training samples increases. Learning curves help evaluate the model's ability to generalise to new, unseen data, and identify issues such as overfitting and underfitting.

The learning curves show the Root-Mean-Square Error (RMSE), representing the deviation of the model prediction from the observation, as a function of the training set size for both the training set (red line) and the test set (blue line). Essentially, a learning curve illustrates experience on the x-axis and learning or improvement on the y-axis. Initially, with small datasets, the training error usually tends to be low as the model can easily fit a few examples. However, with more data added, the training error may increase, leading the model struggles to generalise and fit more complex patterns. At the same time, the error on the test set decreases, indicating that the model becomes more general and can describe the test data. Underfitting happens when both errors are high and close together, suggesting that the model is too simplistic to describe the patterns in the data. On the other hand, overfitting occurs when there is a significant gap between the training and test errors, telling that the model is memorising the training examples and failing to generalise to new data. Convergence between the training set and the test set is expected with large statistics, resulting in stabilized errors for both sets. The learning curves presented in this analysis are reported in Fig. 4.10. It's noted that the first p_T interval has few training samples, but for the rest intervals, the RMSE is stable at set sizes larger than 60000. As expected, the model performance improves with increasing training set size, leading to better description of the data.

Another way to control and validate the model performance is the so-called *AUC-ROC* curve. (AUC: Area Under the Curve). The ROC is the short term for Receiver Operating Characteristics, which is a crucial evaluation metrics for accessing the classification model's performance. Fig. 4.11 shows the curves for the trained model in $2 < p_T < 4 \text{ GeV}/c$ in this analysis, and the rest can be found in Appendix A.2.5. The True Positive Rate (TPR) is shown as a function of the False Positive Rate (FPR). The TPR, defined by Eq. 4.15, represents the fraction of correctly classified instances (true positives) out of all instances (true positives plus false negatives) of the signal class, while the FPR, defined by Eq. 4.16, denotes the fraction of wrongly classified instances (false positives) out of all instances (true positives plus false negatives) of the background class. True and false positives and negatives are potential outcomes when applying a classification to a test set.

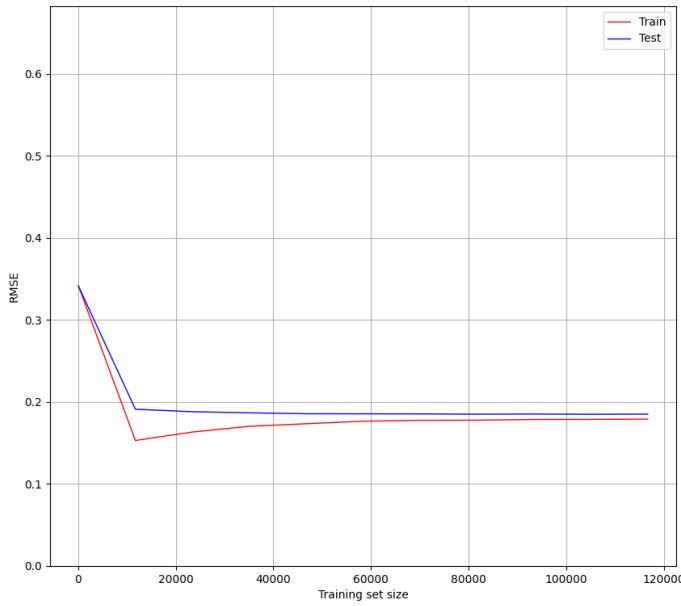


Figure 4.10: The learning curves of the training set and test set in $2 < p_T < 4 \text{ GeV}/c$.

In Fig. 4.11, the blue dashed line shows the ROC curve of the training set, while the solid line describes the test set, in $2 < p_T < 4 \text{ GeV}/c$, and the rest can be found in Fig. A.29 Appendix A.2.5. A large deviation between the two curves means overfitting. The grey dashed line on the diagonal indicates the 50% mark for the test set, where the model prediction is correct in half of the cases. A model with the ROC curve below this line is misinterpreting the data. The AUC can be interpreted as the probability when the model correctly classifies a true positive.

$$TPR = \frac{TP}{TP + FN}, \quad (4.15)$$

$$FPR = \frac{FP}{TP + FN}, \quad (4.16)$$

The resulting model in $2 < p_T < 4 \text{ GeV}/c$ is reported in Fig. A.9, and the rest p_T intervals can be found in Fig. A.30 Appendix A.2.5, showing the scaled signal in red and the background in blue as a function of the BDT output probability for the training and the test set. Furthermore, the importance of the classification criteria is ranked in Fig. 4.13, with the most important features at the top and the less important ones at the bottom. Feature importance refers to the classification that calculates a score for all the input features in a given model, where a higher score means that the variable has a large effect

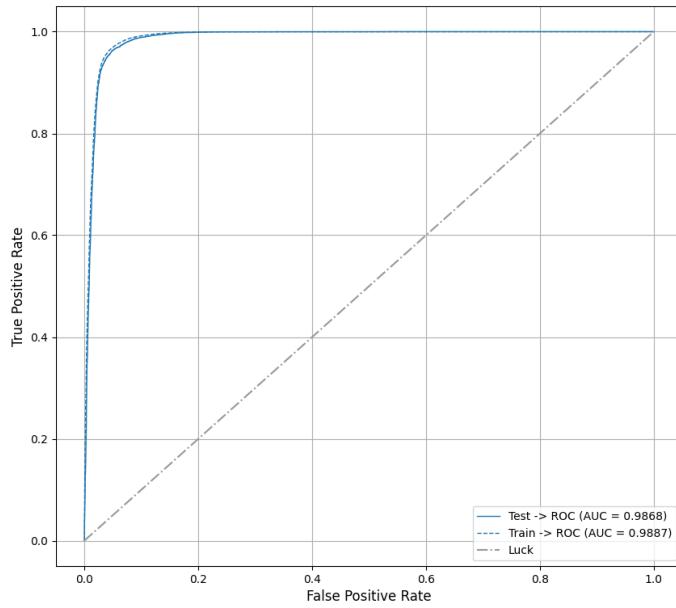


Figure 4.11: The ROC-AUC curves for the training set and test set in $2 < p_T < 4 \text{ GeV}/c$.

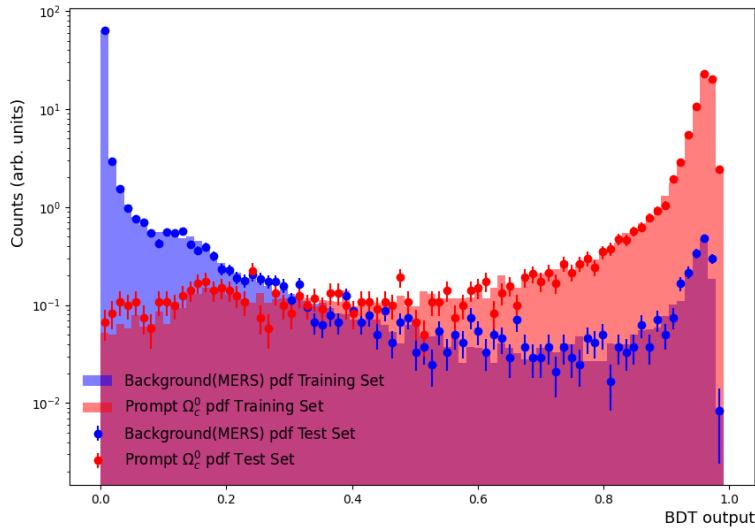


Figure 4.12: Machine learning model in $2 < p_T < 4 \text{ GeV}/c$.

on the model. The other p_T intervals can be found in Fig. A.31 Appendix A.2.5.



博士学位论文
DOCTORAL DISSERTATION

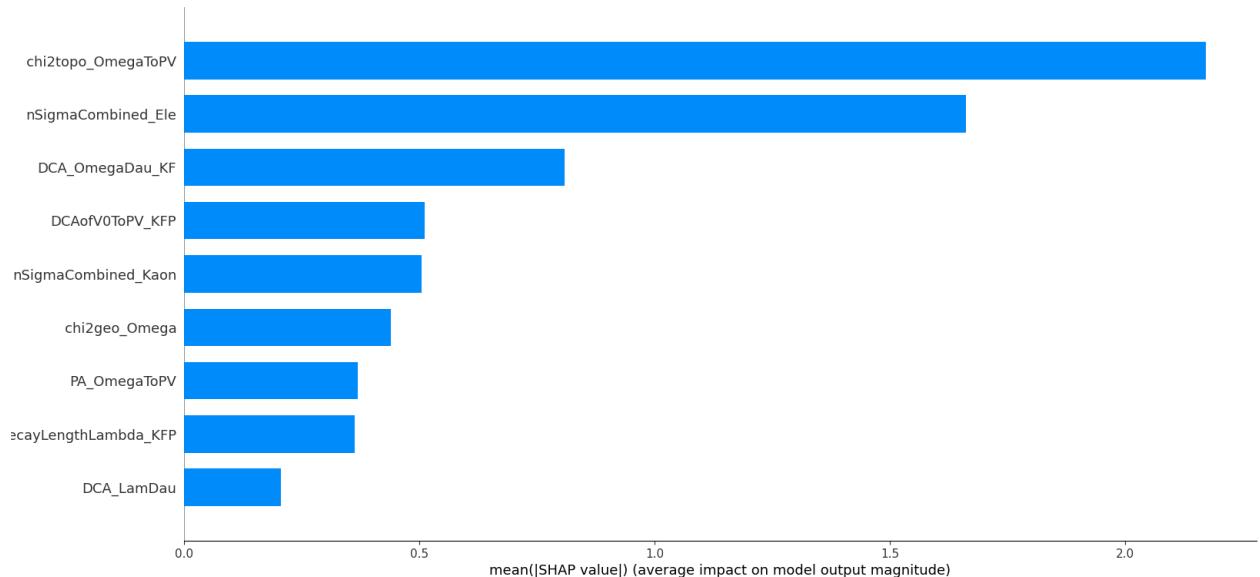


Figure 4.13: Feature importance of the classification criteria in $2 < p_T < 4 \text{ GeV}/c$.



4.3.4 Working point determination

After a model has been trained, it is applied to the full data sample for each p_T interval. Subsequently, the response variable is used as a selection criterion to exclude all candidates below a certain threshold value to reject as many background candidates as possible. If a high probability is selected, the signal purity is large, but at the same time, the efficiency will be also low. To mitigate possible bias from blindly optimising the selection criterion based on the BDT output probability, the *Working Point (WP)* procedure is introduced. The WP is defined by calculating a *pseudo-significance* S , representing the expected significance in data, based on the expected numbers of signal s and the background b candidates, as shown in Eq. 4.17:

$$S = \frac{s}{\sqrt{s+b}} \quad (4.17)$$

The pseudo-significance distribution is computed as a function of the BDT output probability. In this analysis, the background b is estimated from ME(RS) Mixed-Event Right Sign, described above in Section 4.1.3.1. The resulting background distributions are found to be consistent with those from SE, and the SE pairs are used to properly normalise the more statistically abundant ME background samples. The expected signal s , obtained as in Eq. 4.18, is estimated based on the measured production cross section of Ω_c^0 in hadronic decay, assuming that the branching-fraction ratio of $\frac{BR_{semi}}{BR_{hadro}}$, which for Ω_c^0 is the same as the one for Ξ_c^0 [21]. Additionally, the reconstruction efficiency as a function of the BDT selection is determined in this analysis. The efficiency calculation is described later in Section 4.6. The distributions of expected signal and background candidates as a function of BDT probability in $2 < p_T < 4$ GeV/c are shown in Fig. 4.14. The rest p_T intervals can be found in Fig. A.32 and A.33 Appendix A.2.6. The resulting pseudo-significance as a function of the BDT selection in $2 < p_T < 4$ GeV/c, is shown in Fig. 4.15. The rest results can be seen in Fig. A.34 and A.35 Appendix A.2.6.

The corresponding WP values for all p_T intervals are presented in Tab. 4.6.

$$\begin{aligned} N_{\Omega_c^0} &= 2 \cdot \left(\frac{d\sigma}{dp_T dy} \right) \cdot \Delta y \cdot \Delta p_T \cdot \varepsilon_{tot} \cdot \mathcal{L}_{int} \cdot BR_{semi} \\ &= 2 \cdot \left(\frac{d\sigma}{dp_T dy} \cdot BR_{hadro} \right) \cdot \Delta y \cdot \Delta p_T \cdot \varepsilon_{tot} \cdot \mathcal{L}_{int} \cdot \frac{BR_{semi}}{BR_{hadro}}, \end{aligned} \quad (4.18)$$

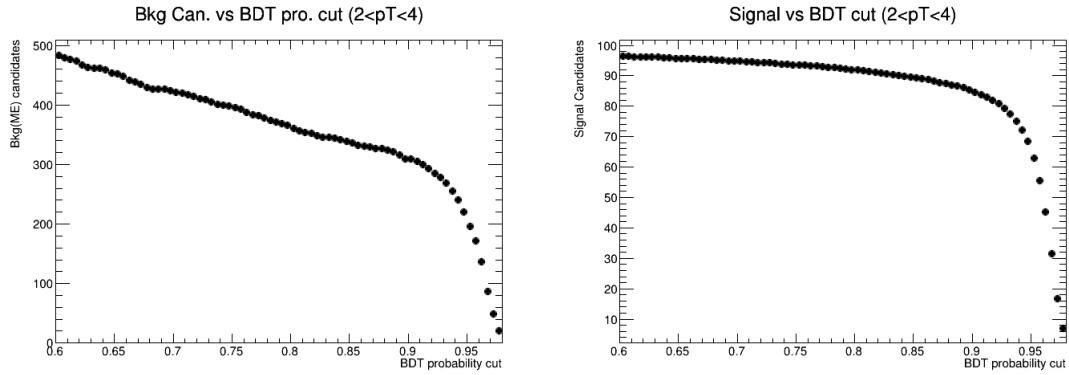


Figure 4.14: Left panel: The number of background candidates (MERS scaled by SEWS) as a function of BDT cut for $2 < p_T < 4$ GeV/c. Right panel: The number of estimated expected signal candidates as a function of BDT cut $2 < p_T < 4$ GeV/c.

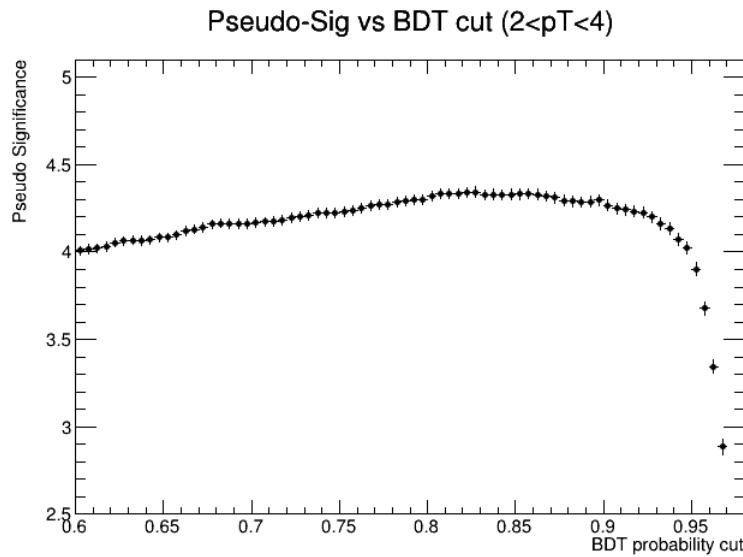


Figure 4.15: The pseudo-significance as a function of BDT cut for $2 < p_T < 4$ GeV/c.

Table 4.6: Determined working point values used for each p_T intervals in this analysis.

p_T (GeV/c)	(1,2)	(2, 4)	(4, 6)	(6, 12)
Working point	0.95	0.86	0.84	0.81



4.4 Signal extraction

After the application of the BDT selection criteria, the invariant-mass distribution of $e\Omega$ pairs can be obtained, as shown in Fig. 4.16. The $e\Omega$ pairs distribution of SEWS is used to normalise the more statistically abundant ME background sample, and compatible background distributions are shown.

To obtain the raw yield, the combinatorial background computed using the ME technique is subtracted from the invariant-mass distribution of $e\Omega$ pairs with opposite-sign charge in SE, as reported in Fig. 4.17. Exploiting a background sample using the SEWS pairs is tested as well, in blue color in Fig. 4.17, and they are compatible with the Cyan color.

The left panel of Fig. 4.18 shows the invariant-mass distribution of $e\Omega$ pairs in SE (same-sign and opposite-sign) and ME (opposite-sign) in the interval $2 < p_T^{e\Omega} < 12 \text{ GeV}/c$. The raw yield is obtained by subtracting the combinatorial background computed using the ME technique from the invariant-mass distribution of $e\Omega$ pairs with opposite-sign charge in the SE. The right panel of Fig. 4.18 shows the invariant-mass distribution of $e\Omega$ candidates, obtained after background subtraction, in comparison with $e\Omega$ opposite-sign charge pairs coming from the Ω_c^0 decay computed with the PYTHIA 8 event

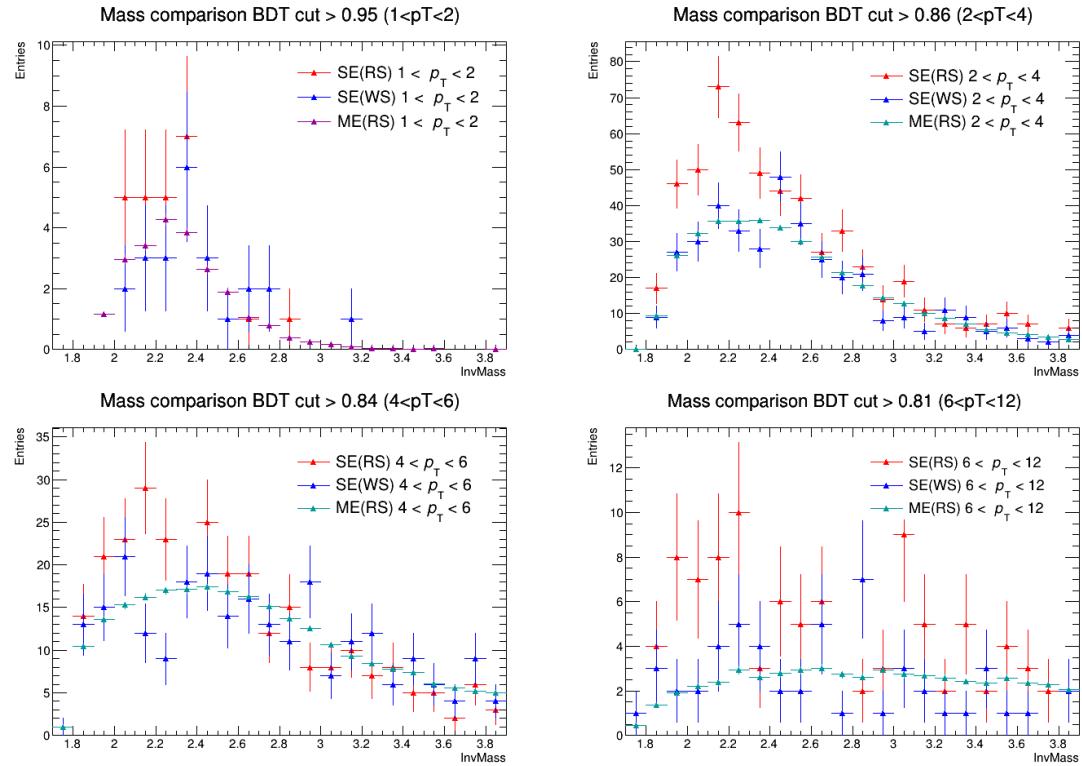


Figure 4.16: The invariant mass distribution of $e\Omega$ for RS and WS in SE, and MERS, at each p_T interval.

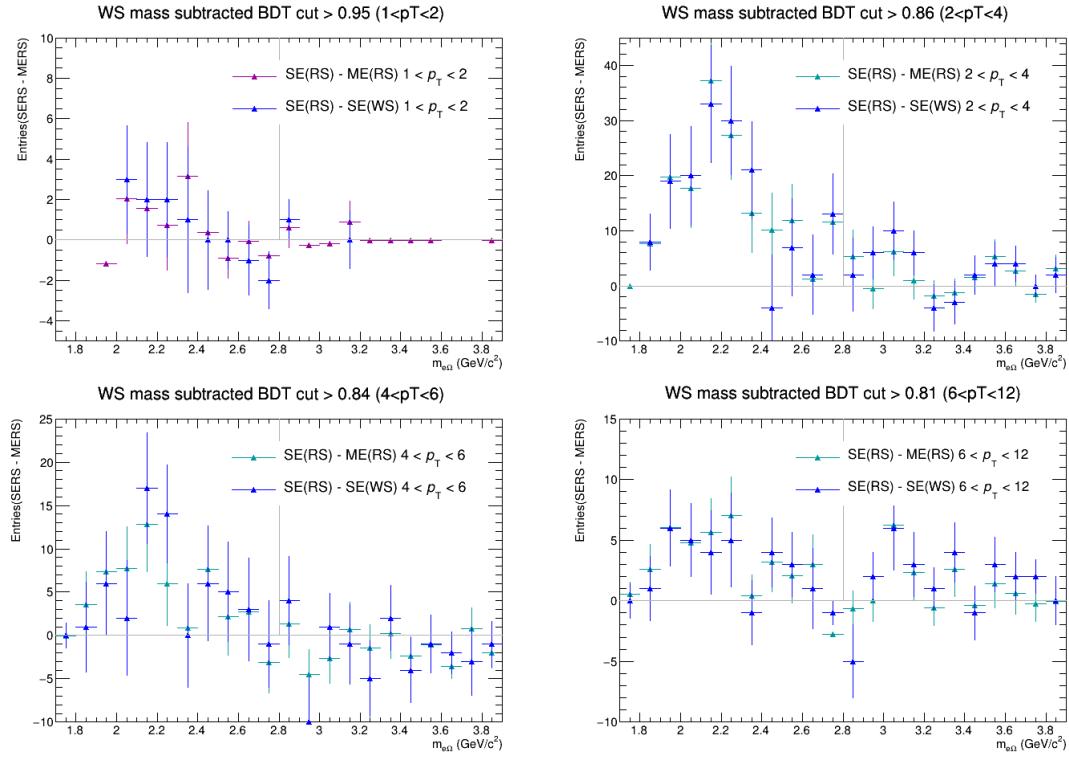


Figure 4.17: The mass distribution comparison: SERS-MERS and SERS-SEWS, at each p_T interval.

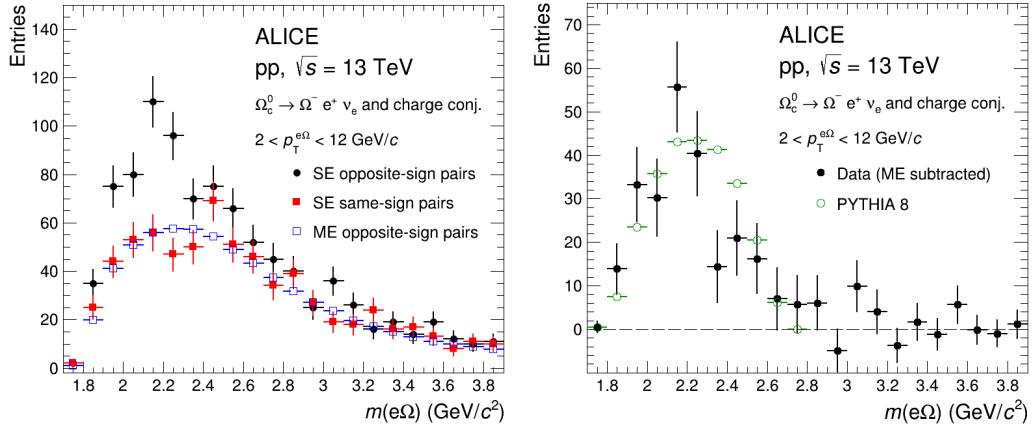


Figure 4.18: Left panel: invariant-mass distribution of opposite-sign pairs (black solid circle marker) and same-sign pairs (red solid square marker) in SE, and opposite-sign pairs (blue open square marker) in ME. Right panel: invariant-mass distribution of the Ω_c^0 candidates obtained by subtracting the opposite-sign charge $e\Omega$ pairs in ME from the opposite-sign charge pairs in SE (black solid circle marker), and $e\Omega$ opposite-sign charge pairs coming from Ω_c^0 decay from PYTHIA 8 (green open circle marker).

generator [13]. Only $e\Omega$ pairs satisfying $1.7 < m_{e\Omega} < 2.7 \text{ GeV}/c^2$ were considered for further analysis.

The number of reconstructed $e\Omega$ signal pairs consists of 232 ± 15 candidates.

As discussed in Section 3.3.1, the prefilter efficiency is used to correct the wrongly tagged electron



博士学位论文
DOCTORAL DISSERTATION

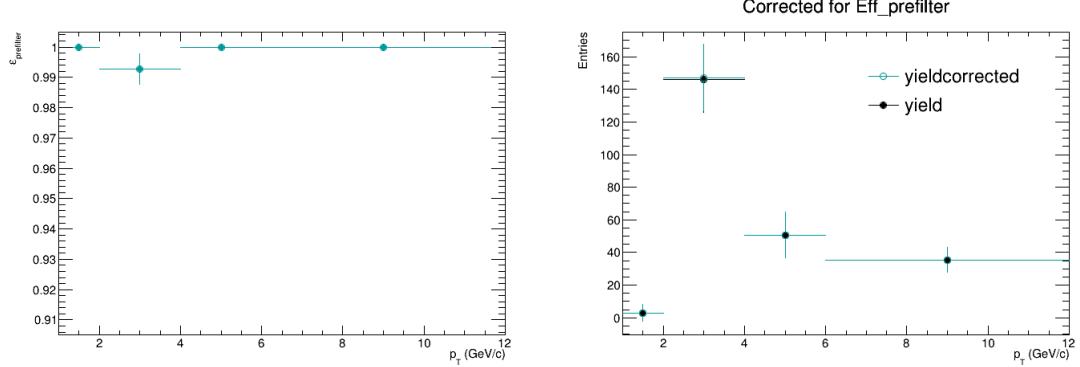


Figure 4.19: Left: The prefilter efficiency. Right: The comparison of yield distribution of p_T $e\Omega$.

candidates. After the correction, the raw yield of candidates shown in Cyan color in the right panel of Fig. 4.19 will be used in the unfolding procedure.



4.5 Unfolding procedure

The unfolding procedure, as introduced previously in Section 3.3.3, corrects for the missing momentum of the neutrino by implementing a response matrix. The Bayesian unfolding technique implemented in the RooUnfold package is employed, requiring two iterations for convergence in this analysis. The response matrix used in the unfolding procedure is defined in the transverse momentum interval $1 < p_T < 12 \text{ GeV}/c$, which is wider than the p_T interval used for the cross-section measurement, to avoid the edge effect at the lowest p_T intervals of the measurement.

The response matrix, representing the correlation between the p_T of the Ω_c^0 baryons and that of the reconstructed e Ω pairs, is shown in the left panel of Fig. 4.20. It is determined with the same simulation setup used for the BDT training.

Furthermore, the weight factor, explained later in Section 4.7.8, is applied to the response matrix. Subsequently, a comparison between the results obtained with and without weights is demonstrated in the right panel of Fig. 4.20. The unfolded spectra with weights (in black) are then corrected for the weighted reconstruction efficiency to get the final result (branching ratio multiplied by cross-section).

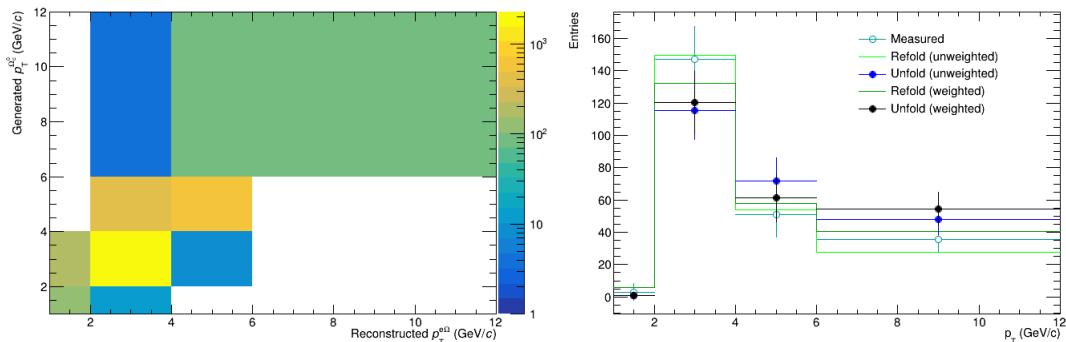


Figure 4.20: Left panel: The correlation matrix between the generated Ω_c^0 -baryon p_T and the reconstructed e Ω pair p_T after weights applied. Right panel: Comparisons of the measured spectrum to the refolded and the unfolded one with and without weights.

4.6 Efficiency correction

Due to the acceptance and inefficiency of the detector, along with the application of prefilter and BDT selections, the extracted signal needs to be corrected for the reconstruction efficiency to compute the final corrected and normalised cross section, as illustrated in Eq. 4.21. The total reconstruction efficiency contains two terms in Eq. 4.19: the product of acceptance times preselection efficiency, and the BDT efficiency (Eq. 4.20).



The preselection efficiency is determined as the ratio of the number of reconstruction candidates after preselection criteria are applied seen Section 4.2, to the total generated candidates. This preselection efficiency is influenced by the acceptance, detector efficiency, and the selection criteria applied after the candidate reconstruction. Notably, the $|y| < 0.8$ cut is applied at both the generated and reconstructed levels. To obtain the BDT efficiency, the ML model is applied to the simulated candidates, and the efficiency is determined by the number of reconstructed candidates passing the BDT selection over the preselected ones.

The Ω_c^0 -baryon p_T distribution from the PYTHIA simulation is reweighted to match the true distribution. The weights are obtained using Tsallis fit to the differential production cross section of Ω_c^0 as measured in Ref. [25], details seen in Section 4.7.8.

Fig. 4.21 shows the detailed efficiency of acceptance and BDT after weights applied. The weighted total reconstruction efficiency of inclusive Ω_c^0 baryons is obtained, as shown in the bottom panel.

$$\varepsilon_{\text{tot}} = (A \times \varepsilon_{\text{preselection}}) \times \varepsilon_{\text{BDT}}, \quad (4.19)$$

$$\varepsilon_{\text{BDT}} = N_{\text{Reco}}^{\Omega_c^0} (\text{with BDT cut}) / N_{\text{Reco}}^{\Omega_c^0} (\text{without BDT cut}), \quad (4.20)$$

The p_T -differential production cross section of inclusive Ω_c^0 baryons multiplied by the BR into the considered semileptonic decay channel is calculated from the yields obtained from the unfolding procedure as follows:

$$\text{BR} \times \frac{d^2\sigma^{\Omega_c^0}}{dp_T dy} = \frac{1}{2\Delta y \Delta p_T} \times \frac{N_{\text{raw}}^{\Omega_c^0}}{(A \times \varepsilon)} \times \frac{1}{\mathcal{L}_{\text{int}}}, \quad (4.21)$$

where $N_{\text{raw}}^{\Omega_c^0}$ is the raw yield (sum of particles and antiparticles) in a given p_T and rapidity interval with width Δp_T and Δy . The factor 1/2 takes into account that the raw yield includes both particles and antiparticles, while the cross section is given for particles only. The \mathcal{L}_{int} is the integrated luminosity. Since the feed-down contribution is not subtracted, the $(A \times \varepsilon)$ factor is the product of detector acceptance and efficiency for inclusive Ω_c^0 baryons, where ε accounts for the reconstruction and selection of the Ω_c^0 decay-product tracks and the Ω_c^0 -candidate selection. The $(A \times \varepsilon)$ correction is obtained from a simulation with the same configuration as the one used for the BDT training and the response matrix. Fig. 4.32 shows the product of the final $(A \times \varepsilon)$ correction factor for inclusive Ω_c^0 as a function of p_T .

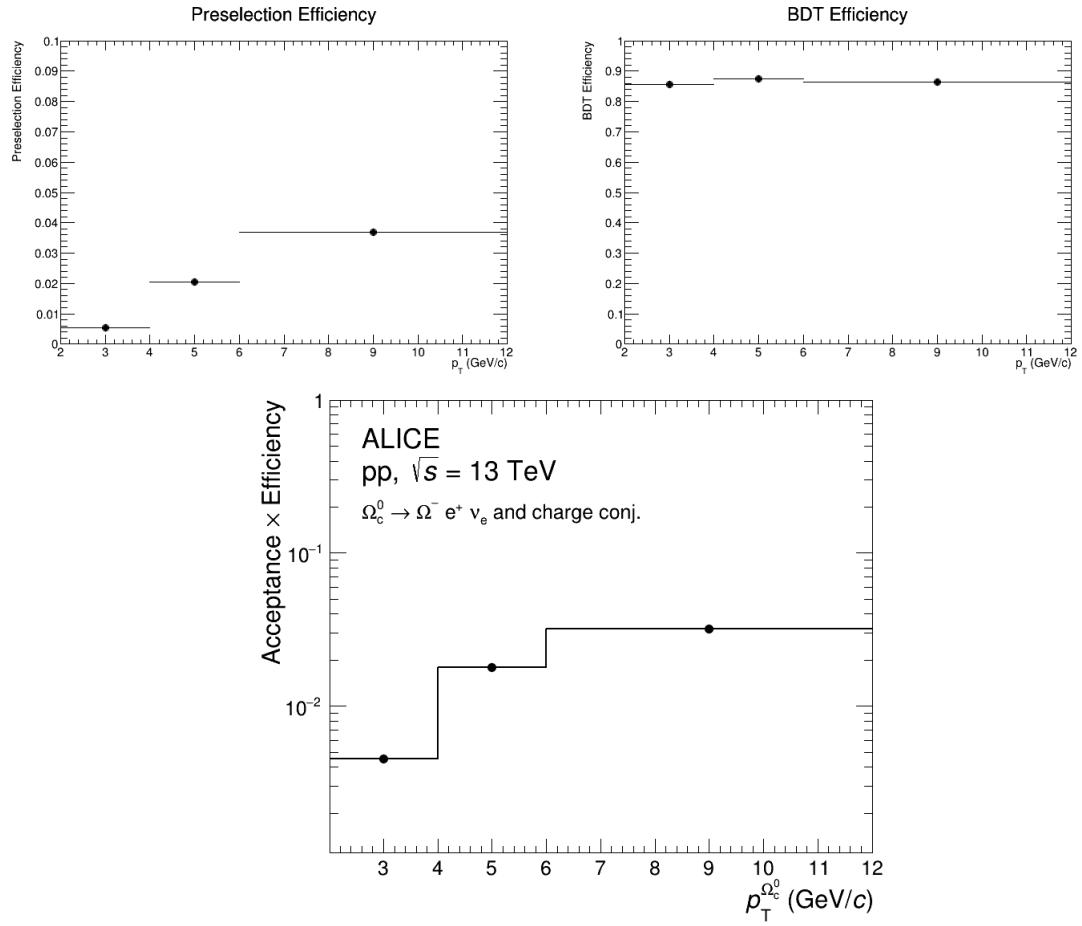


Figure 4.21: Top left panel: (Weighted) Preselection efficiency. Top right panel: (Weighted) BDT efficiency. Bottom panel: (Weighted) Total efficiency.

The comparison of the product of the final ($A \times \varepsilon$) correction factor for inclusive, prompt, and feed-down Ω_c^0 as a function of p_T can be found in Appendix A.2.7.



4.7 Systematic uncertainties

The different contributions to the total systematic uncertainty of the Ω_c^0 production cross section in $2 < p_T < 12 \text{ GeV}/c$ are summarized in Tab. 4.7, which will be explained individually in the following. The various systematic sources are defined as the RMS of the distributions of the corrected yields obtained from the reported variation, if not differently specified.

All systematic uncertainties are considered uncorrelated and summed in quadrature to obtain the total systematic uncertainty. The production cross section has an additional global normalisation uncertainty of 1.6% due to the uncertainties of the integrated luminosity [184]. Its contribution is not summed in quadrature with the other sources of uncertainty in Fig. 4.32 and Fig. 4.33.

Table 4.7: Contributions to the systematic uncertainty of the Ω_c^0 cross section for the p_T intervals $2 < p_T < 4 \text{ GeV}/c$, $4 < p_T < 6 \text{ GeV}/c$, and $6 < p_T < 12 \text{ GeV}/c$.

p_T (GeV/c)	2–4	4–6	6–12
e Ω -pair selection	10%	10%	10%
ITS–TPC matching	2%	2%	2%
Track quality selection	4%	4%	4%
Bayesian-unfolding iterations	4%	4%	4%
Unfolding method	4%	4%	4%
Response-matrix p_T range and binning	10%	10%	–
BDT selection	15%	15%	15%
Generated p_T shape	10%	2%	1%
Total systematic uncertainty	±24%	±22%	±19%
Luminosity		1.6%	

4.7.1 e Ω -pair selection

The systematic uncertainty on the raw-yield extraction is evaluated by investigating possible contaminations to e Ω pairs with opposite charges from other decay channels. The contamination of different decay channels mentioned in Refs. [38, 189] are, i) $\Omega_c^0 \rightarrow \Xi_c^+ e^- \nu_e$, ii) $\Xi_c^0 \rightarrow \Omega^- e^+ \nu_e K^0$, and iii) $\Xi_c^+ \rightarrow \Omega^- e^+ \nu_e K^+$. However, since these decay channels have not been experimentally measured, specific information regarding the particles involved in the possible decay to e Ω pairs is unknown, and a PYTHIA simulation of pp collisions is studied. This simulation is done by using PYTHIA 8.308, a program specifically designed for the generation of high-energy physics, which allows to force of certain



aspects of collision events. Together with the usage of ROOT, version of 6.10/24, a data analysis framework capable of executing codes written in languages like C++ and Python. A premade basic collision event generator code is implemented as a basis on which the rest of the simulation is built, with a specific focus on increasing the production of charm and anti-charm quarks in pp collisions. This approach aims to reduce simulation time, as charm quarks have a low production rate.

Fig. 4.22 shows the results from the PYTHIA 8 simulation study, where the comparison of the invariant-mass distribution of $e\Omega$ pairs between the contaminating decay channel mentioned above and the one in this analysis is illustrated in each p_T interval. It is found that the contributions from the contaminating decay channels are independent of p_T and mainly contribute to a mass region below $2 \text{ GeV}/c^2$.

Fig. 4.23 illustrates how much difference whether the mass cut is applied, giving a 10%. An additional check is done by varying the lower limit of the integration mass range for the signal extraction in the $e\Omega$ mass from 1.7 to 2.2 GeV/c^2 , as shown in Tab. 4.8 and Fig. 4.24, a maximum variation of 10% at the corrected yield level was found, except the last p_T due to large statistical fluctuations.

Considering the two checks, a 10% is considered as the final systematic uncertainty on the raw yield extraction. Note that those decay channels are not experimentally observed, and therefore the Belle [39] and CLEO Collaborations [38] do not correct for it in addition to not assigning a corresponding systematic uncertainty.

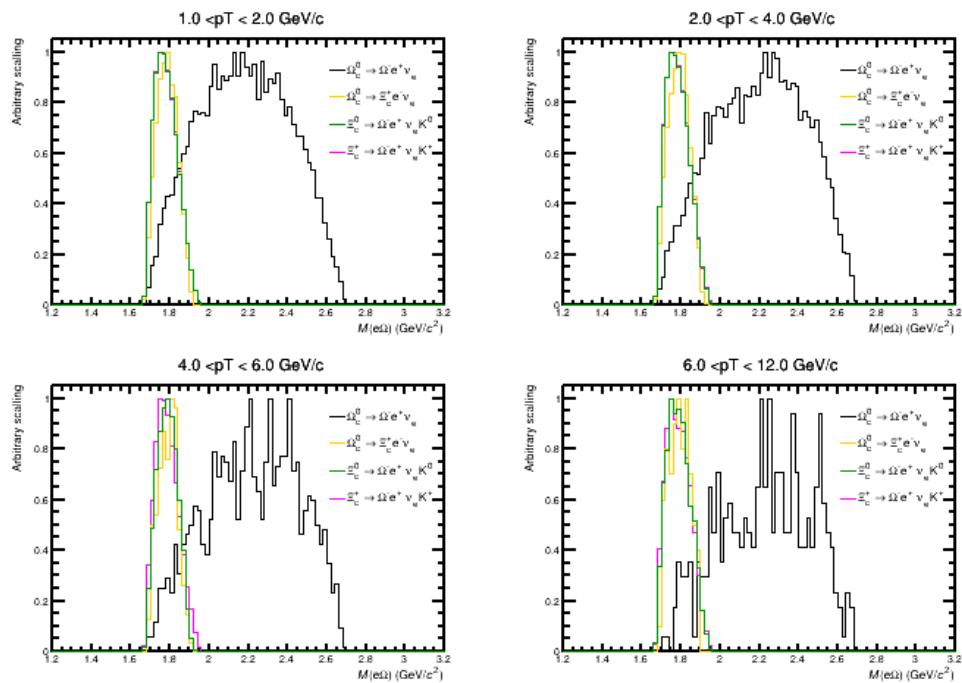


Figure 4.22: Invariant-mass distribution of the $e\Omega$ pairs from PYTHIA 8 simulation study in different p_T intervals.

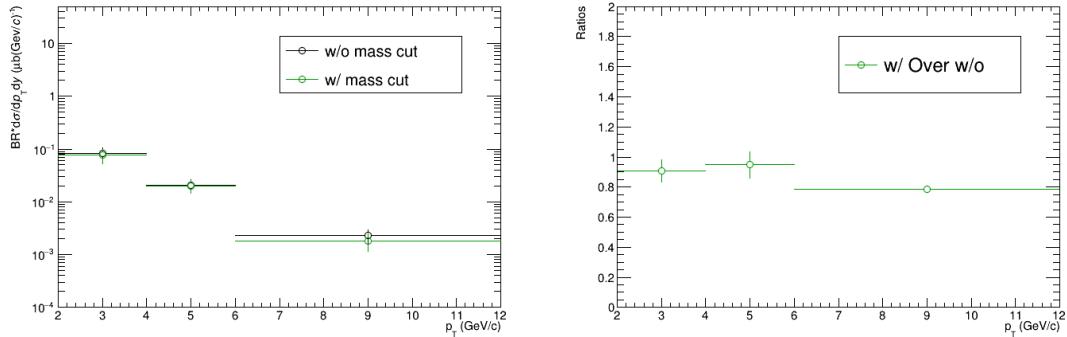


Figure 4.23: Left panel: Comparison of the Ω_c^0 spectrum with and without the mass cut ($M_{e\Omega} > 2\text{GeV}/c^2$) applied. Right panel: the ratio between the Ω_c^0 spectrum with mass cut over Ω_c^0 spectrum without one.

Table 4.8: The cut variation on $M_{e\Omega}$ (GeV/c^2).

Cut variable	VeryLoose1	VeryLoose	Loose	Standard	Tight
$M_{e\Omega}$ (GeV/c^2)	1.8	1.9	2	2.1	2.2

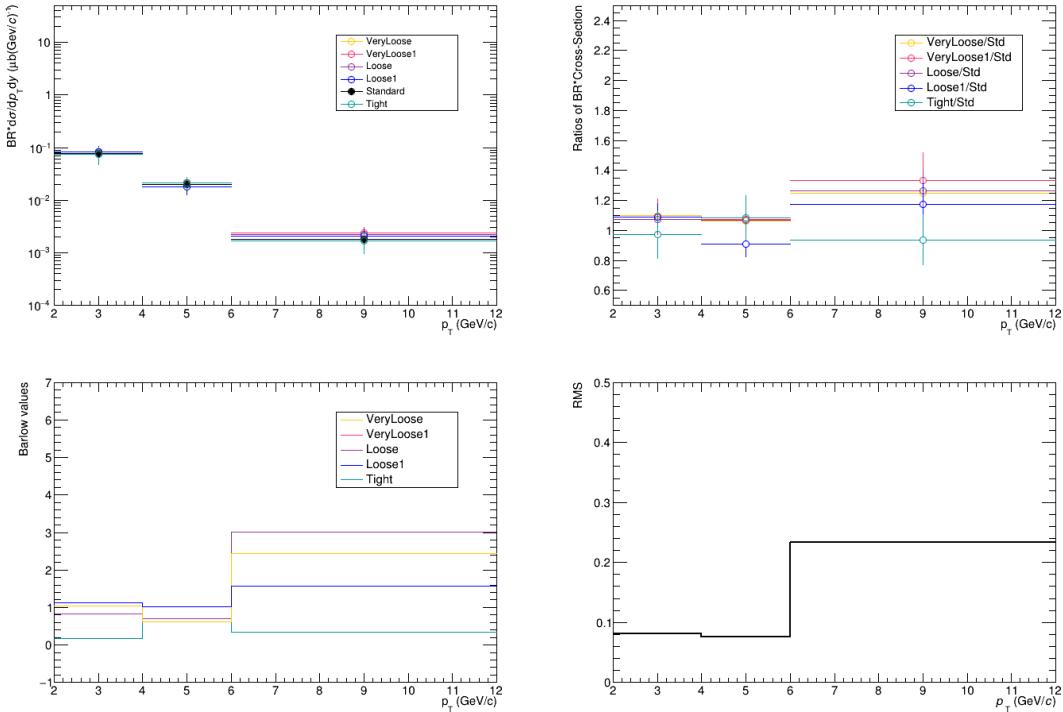


Figure 4.24: Top left panel: the Ω_c^0 spectrum with different mass selections. Top right panel: ratios of the Ω_c^0 spectrum with variations. Bottom left panel: Barlow value. Bottom right panel: the root mean square (RMS) of the deviations from 1.



4.7.2 ITS–TPC matching

As discussed in Section 3.5.1.3, one of the systematic uncertainties on the tracking efficiency is determined by comparing the matching efficiency of prolonging a track from the TPC to the ITS in data and simulation. The uncertainties on electron tracks are propagated to the Ω_c^0 candidates according to the decay kinematics, resulting in an uncertainty of 2%, as shown in Tab. A.10. More details can be seen Appendix A.2.8.

Table 4.9: The systematic uncertainty of $p_T^{\Omega_c^0}$ propagated from p_T^e .

$p_T^{\Omega_c^0}$ (GeV/c)	2-4	4-6	6-12
Systematic	2%	2%	2%

4.7.3 Track quality selection

The second contribution to the track reconstruction is estimated by varying the track quality selection, seen Tab. 4.10. Those varied track selection cuts will be implemented at the pre-selection step.

Fig. 4.25 shows the detailed comparisons, and 4% is assigned for this systematic study.

Table 4.10: The variations of TPC selection criteria in the pre-selection used for this systematic study.

Cuts variables	VeryLoose	Loose	Standard	Tight	VeryTight
Number of TPC crossed rows	> 65	> 70	> 70	> 75	> 80
TPC crossed rows/findable ratio	> 0.75	> 0.75	> 0.80	> 0.85	> 0.90
Number of TPC clusters for dE/dx	> 40	> 45	> 50	> 55	> 60

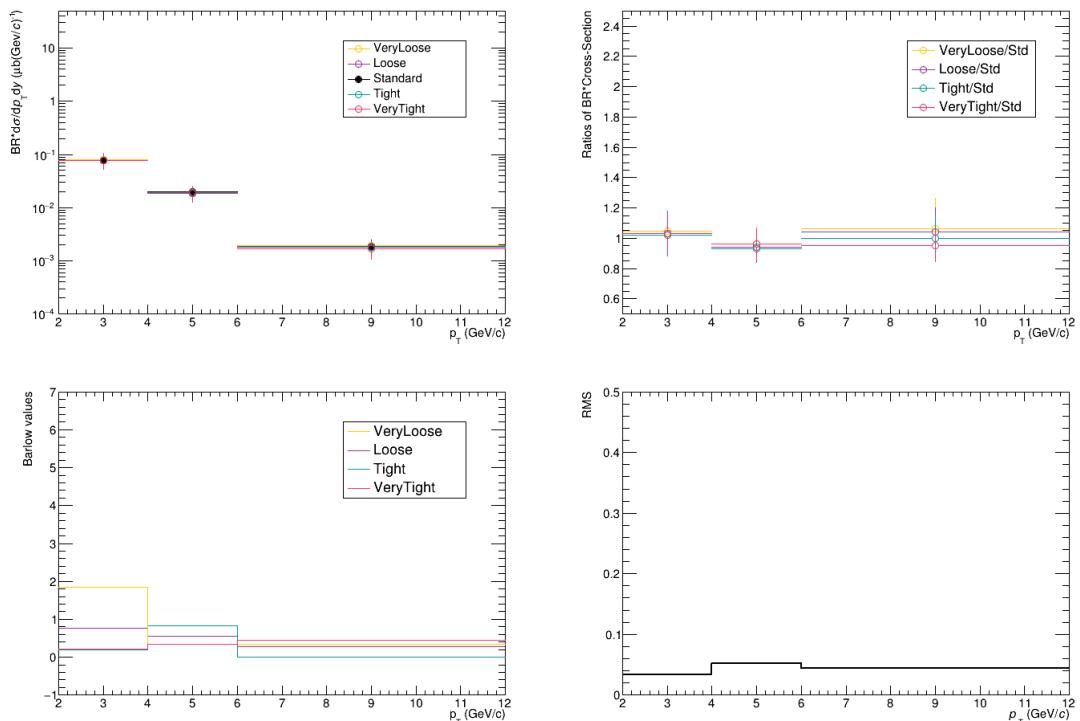


Figure 4.25: Top left panel: The Ω_c^0 spectrum with different track selection criteria. Top right panel: The ratios. Bottom left panel: The Barlow value for each variation. Bottom right panel: The RMS of the deviation.



4.7.4 Bayesian-unfolding iterations

The first contribution to the systematic uncertainty of the unfolding procedure is due to the regularisation procedure in the Bayesian unfolding. It was estimated by varying the iteration number between 2 and 5. There is 4% systematic uncertainty, details are shown in Fig. 4.26.

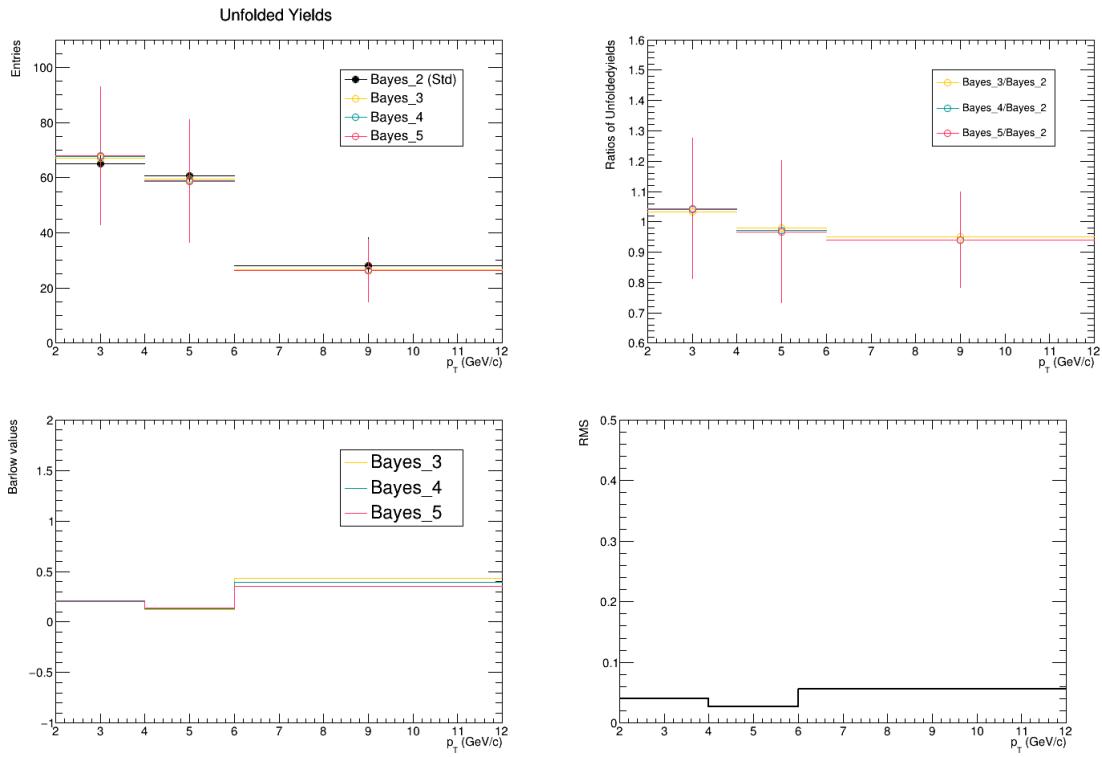


Figure 4.26: Top left panel: the comparisons of the distribution of unfolded yields with the different variations. Top right panel: the ratios of the spectra with different unfolding methods relative to the spectrum with the standard procedure. Bottom left panel: the Barlow values for each variation. Bottom right panel: the rms values of the deviation from 1.



4.7.5 Unfolding method

The second contribution to the systematic uncertainty of the unfolding procedure comes from the application of a different unfolding method, namely the Singular value decomposition (SVD) [159]. The SVD method is sensitive to the d vector, which is obtained using the TSVDUnfoldlocal::GetD() function. Overregularisation (i.e., choosing two small a value for k) will bias the unfolded spectra towards the MC input, while underregularisation (i.e., choosing two large a value for k) will lead to large fluctuations. The number of regularisations is strongly dependent on the bins used in the analysis. In this study, only four bins are employed for the unfolding procedure, resulting in the d vector being determined at four values. The d vector distribution can be observed in Fig. 4.27. Ideally, the magnitude of $|d|$ should be close to unity, suggesting that $k = 3$ and $k = 4$ are suitable choices for the systematic checks to avoid significant fluctuations.

The comparisons can be found in Fig. 4.28, where a 4% dependence is assigned for the systematic uncertainty.

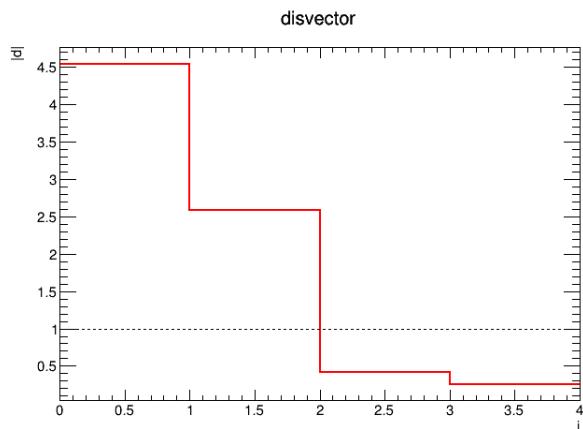


Figure 4.27: The distribution of d vector.

4.7.6 Response-matrix p_T range and binning

The third source to the systematic uncertainty of the unfolding procedure is related to the sensitivity of the unfolding to bin edge effects used in response matrix. An uncertainty of 10% is assigned in the interval $2 < p_T < 6 \text{ GeV}/c$, as shown in Fig. 4.29. At high p_T , this systematic uncertainty is negligible, as the results are converged with larger iteration nubmer.

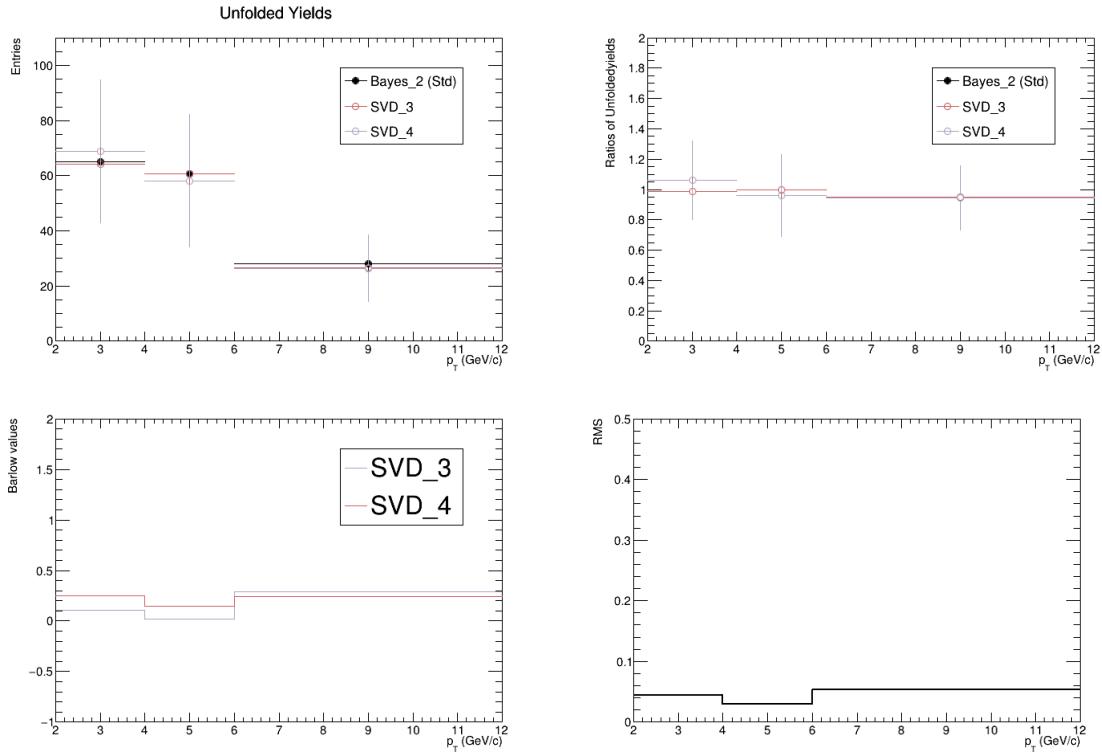


Figure 4.28: Top left panel: the comparisons of the distribution of unfolded yields with the different variations. Top right panel: the ratios of the spectra with different unfolding methods relative to the spectrum with the standard procedure. Bottom left panel: the Barlow values for each variation. Bottom right panel: the rms values of the deviation from 1.

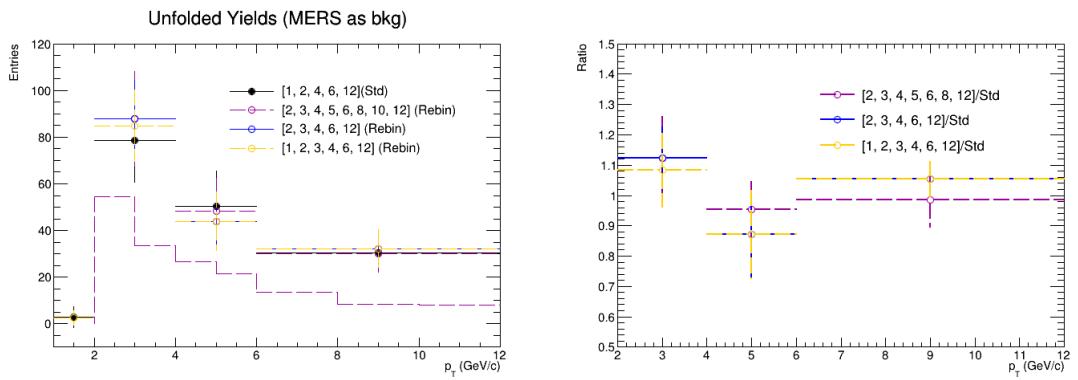


Figure 4.29: Left panel: the comparisons of the distribution of unfolded yields with the different p_T intervals. Right panel: ratios of the varied distributions with different p_T intervals to the default one.



4.7.7 BDT selection

The systematic uncertainty on the selection efficiency can be from the imperfections in the description of the detector response and alignment in the simulation. The systematic uncertainty is estimated by varying the BDT probability selection and recomputing the cross section for each BDT selection, which results in the modification of the efficiency, raw yield, and background values. If the simulations accurately describe the real data, the uncertainty will be minimal, as the calculated efficiency accounts for losses.

To ensure signal stability and avoid statistical fluctuations, the systematic evaluation is extended using a BDT model with different training variables (excluding PID information) and preselection criteria (PID selection is varied when not included in the BDT). Specifically:

- Estimate systematic uncertainty by exploiting a new BDT model in which electron PID information is removed from training and the $n\sigma_{\text{TPC}}^e$ which is p_T dependent, added as a preselection criterion. 10 loose and 10 tight BDT selections are chosen based on the default one. In Appendix A.2.9, Fig. A.39, Fig. A.40 and Fig. A.41 are showing the detailed comparison for the three p_T intervals. The resulting systematic uncertainty is obtained by adding in quadrature the root mean square (RMS) and the shift of the distribution with respect to the central value: 16% in $2 < p_T < 4$, 17% in $4 < p_T < 6$, and 13% in $6 < p_T < 12$.
- Five variations of the p_T -dependent $n\sigma_{\text{TPC}}^e$ preselection are examined, as listed in Appendix A.2.9 Tab. A.11. The detailed comparison is seen in Fig. 4.30. Except for the last p_T bin, where significant fluctuations are observed, a 10% uncertainty is assigned for this check.

Combined with the two checks, 15% is assigned for the BDT selections for the three p_T intervals.

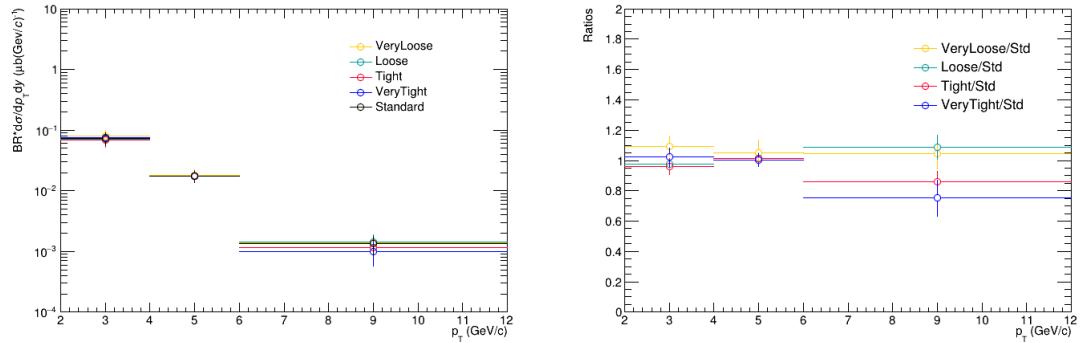


Figure 4.30: Left panel: the comparison of the Ω_c^0 spectrum with variations of the p_T dependent $n\sigma_{\text{TPC}}^e$. Right panel: the ratios of the Ω_c^0 spectrum with variations.



4.7.8 Generated p_T shape

The systematic uncertainty on the efficiency, due to the shape of the true and generated Ω_c^0 p_T distributions, is estimated by varying the p_T weights titling the Tsallis fit within the statistical and p_T uncorrelated uncertainties of the differential production cross section of Ω_c^0 reported in Ref. [25].

In Fig. 4.31, different fitting procedure shown in green, orange, magenta color as in the top panel. The resulting three weight factors are illustrated in the top right panel. In the bottom left panel, total acceptance efficiency with different weights factors applied and without weight is reported. The assigned uncertainty, defined as the maximum variation observed, is 10% in the interval $2 < p_T < 4 \text{ GeV}/c$, 2% in the interval $4 < p_T < 6 \text{ GeV}/c$, and 1% for highest p_T interval, as shown in the bottom right panel.

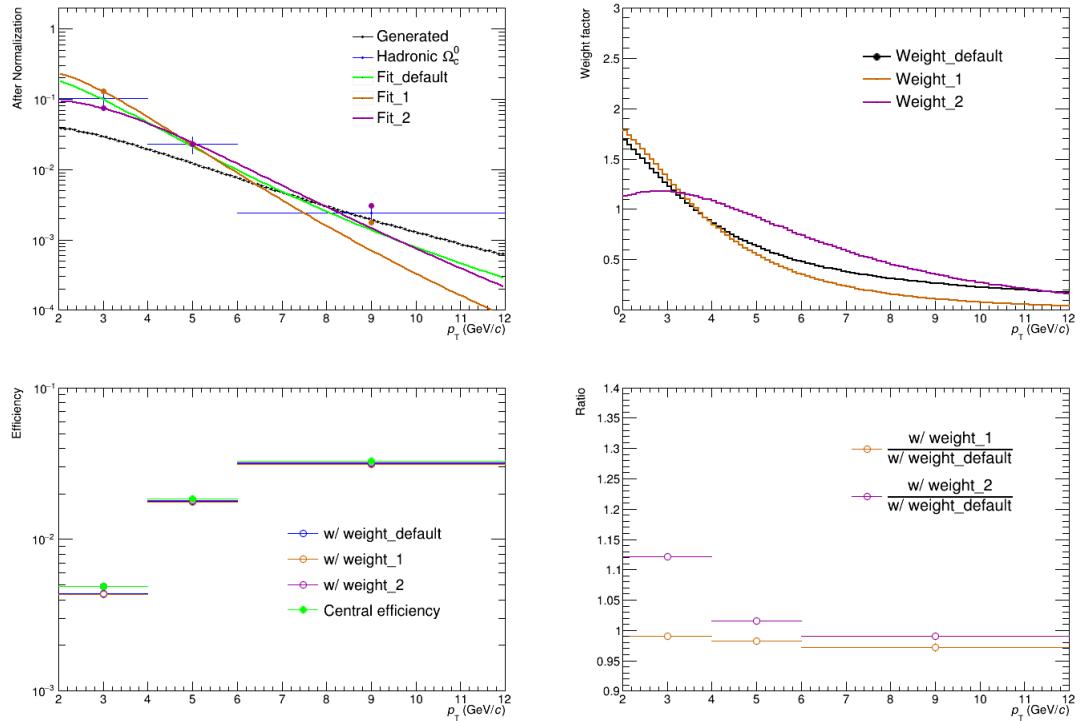


Figure 4.31: Top left panel: the Ω_c^0 spectrum distribution between the Generated and hadronic result [25], and also three different Tsallis fitting to the hadronic result. Top right panel: the three weighting functions. Bottom left panel: comparisons of the reconstruction efficiency after re-weighting procedure with three different weighting functions. Bottom right panel: ratios of the reconstruction efficiency efficiency between the varied Tsallis fitting functions to the central one.



4.8 Results

The p_T -differential cross section of inclusive Ω_c^0 baryon production multiplied by the branching ratio into $\Omega^- e^+ \nu_e$, in pp collision at $\sqrt{s} = 13$ TeV, measured in rapidity interval $|y| < 0.8$ and the p_T interval $2 < p_T < 12$ GeV/c, is shown in the top panel of Fig. 4.32. It is compared with the previously published measurements of inclusive Ω_c^0 baryon production in the hadronic decay channel $\Omega^- \pi^+$. The error bars and boxes represent the statistical and systematic uncertainty, respectively. The uncertainty of the integrated luminosity is not included in the boxes. The bottom panel of Fig. 4.32 reports the branching-fraction ratio $\text{BR}(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e) / \text{BR}(\Omega_c^0 \rightarrow \Omega^- \pi^+)$ as a function of p_T . The systematic uncertainty on the branching-fraction ratio is calculated assuming all the uncertainties between the two measurements as uncorrected, except for the ITS–TPC matching efficiency, track quality selection, and the MC p_T shape. The uncertainty of the luminosity cancels in the ratio, as it is fully correlated.

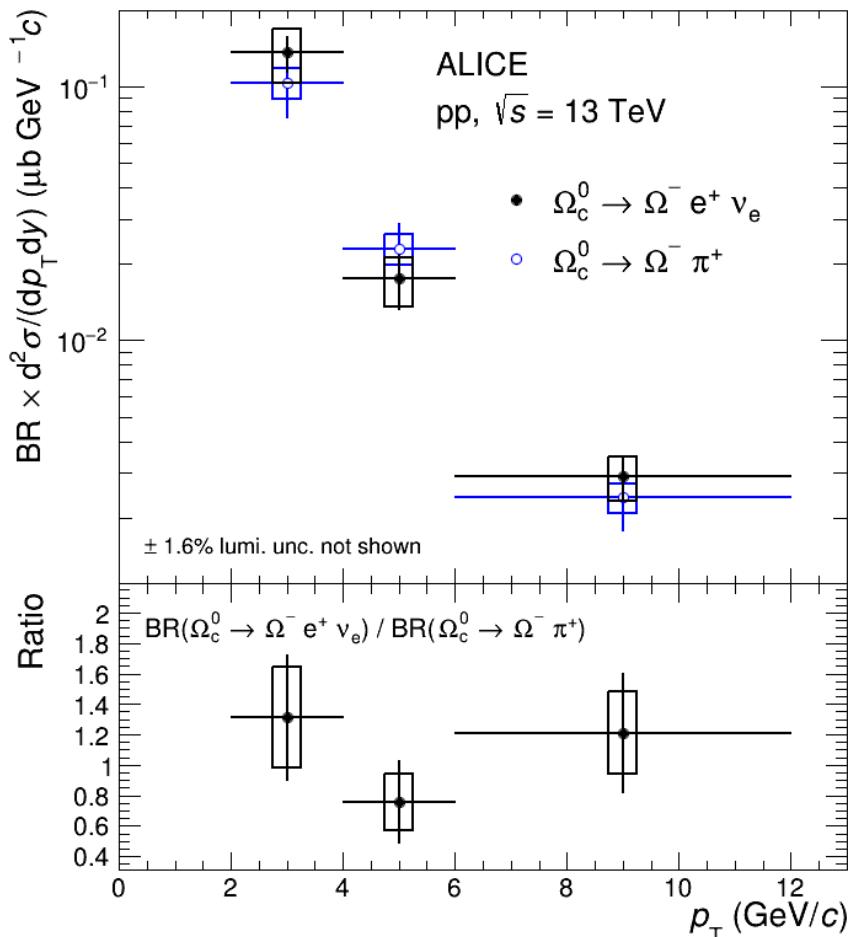


Figure 4.32: Top panel: p_T -differential production cross sections of inclusive Ω_c^0 baryons multiplied by the branching ratios into $\Omega^- e^+ \nu_e$ and $\Omega^- \pi^+$ [25] in pp collisions at $\sqrt{s} = 13$ TeV. Bottom panel: p_T -differential branching-fraction ratio $\text{BR}(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e) / \text{BR}(\Omega_c^0 \rightarrow \Omega^- \pi^+)$.



The ratio of the two measurements shown in the bottom panel of Fig. 4.33, is used to calculate the p_T independent branching-fraction ratio $\text{BR}(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e)/\text{BR}(\Omega_c^0 \rightarrow \Omega^- \pi^+)$. The results are averaged over p_T using the inverse uncorrelated relative uncertainties as weights [190]. The weights are defined as the sum in quadrature of the relative statistical uncertainties and the p_T -uncorrelated part of the systematic uncertainties. All the systematic uncertainties are considered as p_T correlated in the semileptonic decay. For the hadronic decay, all systematic uncertainties are considered as p_T correlated, except for the raw yield extraction. The p_T -correlated systematic uncertainties are propagated by recomputing the ratio after shifting up and down the ratios with the corresponding p_T -correlated systematic uncertainties, as shown in Eq. 4.27 and Eq. 4.28. The detailed calculation procedure is illustrated from Eq. 4.22 to Eq. 4.28. The final systematic uncertainty on the ratio is obtained by summing the p_T -correlated and uncorrelated systematic uncertainties in quadrature.

In Fig. 4.33, the measured ratio is $\text{BR}(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e)/\text{BR}(\Omega_c^0 \rightarrow \Omega^- \pi^+) = 1.12 \pm 0.22 \text{ (stat.)} \pm 0.27 \text{ (syst.)}$. The branching-fraction ratio is compared with previous experimental measurements: $2.4 \pm 1.1 \text{ (stat.)} \pm 0.2 \text{ (syst.)}$ from the CLEO Collaboration [38], and $1.98 \pm 0.13 \text{ (stat.)} \pm 0.08 \text{ (syst.)}$ from the Belle Collaboration [39], and with the theory predictions based on the light-front approach and light-cone sum rules calculations [40, 41]: 1.1 ± 0.2 and 0.71 , also 1.35 is computed by applying the effective colour approach under the framework of $SU(3)_f$ symmetry [42]. The ALICE result is compatible within 1σ with the CLEO result and it is 2.3σ lower than the one measured by the Belle Collaboration. The ALICE measurement is also consistent within 1σ with the available theoretical predictions, which showed some tensions when compared to the Belle results. The present result is also compatible within the uncertainties with the $\text{BR}(\Xi_c^0 \rightarrow \Xi^- e^+ \nu_e)/\text{BR}(\Xi_c^0 \rightarrow \Xi^- \pi^+)$ measured by the ALICE Collaboration [21]. The agreement between those two measurements is also predicted by the light-front approach calculations [40, 191]. In view of those future measurements, it would be beneficial to compare also with additional model calculations, like lQCD [192] and QRM [193] which already provide their predictions for the branching-fraction ratio $\text{BR}(\Xi_c^0 \rightarrow \Xi^- e^+ \nu_e)/\text{BR}(\Xi_c^0 \rightarrow \Xi^- \pi^+)$.

$$w_i = 1/\sqrt{\left(\frac{\sigma_i^{\text{stat}}}{R_i}\right)^2 + \left(\frac{\sigma_i^{\text{pTuncorr}}}{Y_i}\right)^2}, \quad (4.22)$$

$$\langle R_i \rangle_{p_T} = \frac{\sum_i \left(R_i \cdot \frac{1}{w_i^2}\right)}{\sum_i \left(\frac{1}{w_i^2}\right)}, \quad (4.23)$$



$$\langle \sigma_{\text{stat}} \rangle_{p_{\text{T}}} = \frac{\sqrt{\sum_i \left(\sigma_i^{\text{stat}} \cdot \frac{1}{w_i^2} \right)^2}}{\sum_i \left(\frac{1}{w_i^2} \right)}, \quad (4.24)$$

$$\langle \sigma_{\text{syst}}^{\text{pTuncorr}} \rangle_{p_{\text{T}}} = \frac{\sqrt{\sum_i \left(\sigma_i^{\text{pTuncorr}} \cdot \frac{1}{w_i^2} \right)^2}}{\sum_i \left(\frac{1}{w_i^2} \right)}, \quad (4.25)$$

$$\langle \sigma_{\text{syst}}^{\text{pTcorr_up}} \rangle_{p_{\text{T}}} = \langle R_i \rangle_{p_{\text{T}}} \sqrt{\left(\frac{\langle R^{\text{pTcorr_down; hadro}} \rangle_{p_{\text{T}}} - 1}{\langle R_i \rangle_{p_{\text{T}}} - 1} \right)^2 + \left(\frac{\langle R^{\text{pTcorr_up; semi}} \rangle_{p_{\text{T}}} - 1}{\langle R_i \rangle_{p_{\text{T}}} - 1} \right)^2}, \quad (4.26)$$

$$\langle \sigma_{\text{syst}}^{\text{pTcorr_down}} \rangle_{p_{\text{T}}} = \langle R_i \rangle_{p_{\text{T}}} \sqrt{\left(\frac{\langle R^{\text{pTcorr_up; hadro}} \rangle_{p_{\text{T}}} - 1}{\langle R_i \rangle_{p_{\text{T}}} - 1} \right)^2 + \left(\frac{\langle R^{\text{pTcorr_down; semi}} \rangle_{p_{\text{T}}} - 1}{\langle R_i \rangle_{p_{\text{T}}} - 1} \right)^2}, \quad (4.27)$$

$$\langle \sigma_{\text{syst}}^{\text{pTcorr}} \rangle_{p_{\text{T}}} = \max[\langle \sigma_{\text{syst}}^{\text{pTcorr_up}} \rangle_{p_{\text{T}}}, \langle \sigma_{\text{syst}}^{\text{pTcorr_down}} \rangle_{p_{\text{T}}}], \quad (4.28)$$

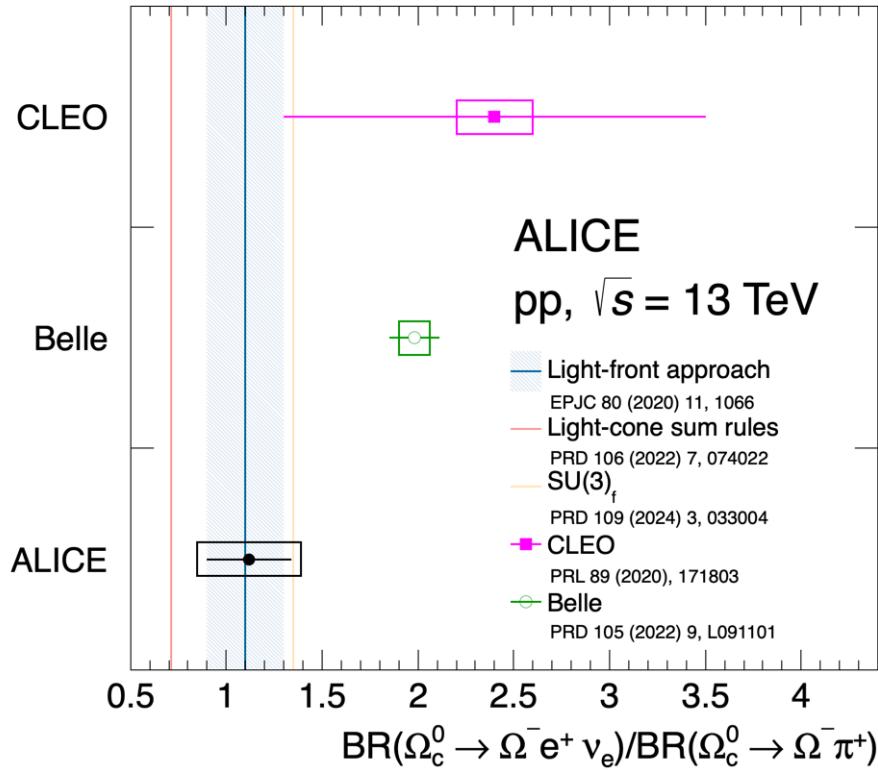


Figure 4.33: Comparison of $\text{BR}(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e) / \text{BR}(\Omega_c^0 \rightarrow \Omega^- \pi^+)$ between experimental and theoretical calculations [38–42].



4.9 Summary

The inclusive p_T -differential production cross section of the charm-baryon Ω_c^0 multiplied by the branching ratio $\text{BR}(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e)$ is measured for the first time at midrapidity ($|y| < 0.8$), in the transverse-momentum interval $2 < p_T < 12 \text{ GeV}/c$, in pp collisions at $\sqrt{s} = 13 \text{ TeV}$. To improve the precision of the measurement, new techniques are employed in this analysis, as explained in Section 4.1. Appendix A.2.2, and Appendix A.2.3 illustrate the advantages of using these new methods, which help in the raw yield extraction reported in Fig. A.19, as well as improve the relative statistics error of corrected yield seen in Fig A.21 and Tab. A.7.

The result is compared with the measurement of the Ω_c^0 -baryon cross section in its hadronic decay $\Omega_c^0 \rightarrow \Omega^- \pi^+$ [25]. The $\text{BR}(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e)/\text{BR}(\Omega_c^0 \rightarrow \Omega^- \pi^+)$ is measured to be $1.12 \pm 0.22 \text{ (stat.)} \pm 0.27 \text{ (syst.)}$, using the inclusive production cross section measured in Ref. [25]. The branching-fraction ratio is consistent within 1σ with the CLEO result, and 2.3σ lower than the value reported by the Belle Collaboration. The value is also in good agreement with the available theoretical predictions.



5 Summary and outlook

5.1 Summary

The research presented in this dissertation contributes significantly to the understanding of charm-baryon production in small systems at the LHC, as well as the charm quark hadronisation mechanism. Specifically, the two analyses focus on the charm-strange baryons, Ξ_c^0 and Ω_c^0 , measured at midrapidity in pp collisions. Chapter 3 reports the measurement of prompt Ξ_c^0 at $\sqrt{s} = 5.02$ TeV, while Chapter 4 details the measurement of inclusive Ω_c^0 at $\sqrt{s} = 13$ TeV. Both baryon species are reconstructed via their semileptonic decay channels, $\Xi_c^0 \rightarrow \Xi^- e^+ \nu_e \rightarrow (\Lambda \pi^-) e^+ \nu_e \rightarrow ((\pi^- p) \pi^-) e^+ \nu_e$ and $\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e \rightarrow (\Lambda K^-) e^+ \nu_e \rightarrow ((\pi^- p) K^-) e^+ \nu_e$, respectively, along with their charge conjugates.

The production of prompt Ξ_c^0 is measured within the transverse momentum interval of 2 to 8 GeV/c . The ratio of the p_T -differential Ξ_c^0 -baryon and D^0 -meson production cross sections is also presented. These two results are compared with calculations of theoretical models implementing different particle production and hadronisation mechanisms, revealing interesting insights into charm-production mechanisms. The experimental result is systematically underestimated by the PYTHIA 8 event generator with the Monash 2013 tune, which uses string fragmentation tuned on e^+e^- data. It is also underestimated when PYTHIA8 is tuned with colour reconnection beyond the leading-colour approximation, which introduces additional junction topologies favouring baryon production. Both of them systematically underestimate the experimental result. The latter predicts an enhanced production of baryons and is closer to the data compare to PYTHIA 8 Monash 2013. The quark (re-)combination model (QCM) underpredicts the Ξ_c^0/D^0 yield ratio by the same amount as it does for Ξ_c^0 production, namely by a factor 2–3 for $p_T < 4 \text{ GeV}/c$. The p_T -differential Ξ_c^0/D^0 yield ratio is also compared with a prediction from the statistical hadronisation model (SHM) with feed down from an augmented set of higher-mass charm baryon states calculated by the relativistic quark model (RQM), which gives lower values than the measured data. By contrast, the CR model and the SHM model (including RQM) both describe the Λ_c^+/D^0 yield ratio reasonably well. The Catania model, which assumes that the charm-quark hadronisation processes happen via fragmentation and coalescence, is the one that best captures the Ξ_c^0/D^0 yield ratio in the full p_T interval. Catania provides a possible new scenario in pp collisions at LHC energies, assuming that low- p_T charm quarks hadronise via a combination of fragmentation and additional coalescence mechanisms. Besides that, the result of Ξ_c^0 reported in this thesis provides additional information on the non-universality of charm fragmentation from e^+e^- and ep collisions to pp collisions.



The inclusive production of the charm-strange baryon Ω_c^0 is measured for the first time via its semileptonic decay into $\Omega^- e^+ \nu_e$ with the ALICE detector. In order to improve the precision of these measurements, three additional innovative techniques are introduced: 1) the use of the Kalman-Filter vertexing algorithm to reconstruct the Ω cascade; 2) the implementation of a machine learning approach based on the Boosted Decision Tree algorithm to enhance the signal extraction performance; 3) the adoption of the mixed-event technique for background samples to help the machine learning training and signal extraction. The p_T -differential cross section of inclusive Ω_c^0 -baryon production multiplied by the branching ratio into $\Omega^- e^+ \nu_e$ is compared with previously published results of inclusive Ω_c^0 -baryon production in the hadronic decay channel $\Omega^- \pi^+$ measured in the same p_T interval. The p_T -differential branching-fraction ratio $BR(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e)/BR(\Omega_c^0 \rightarrow \Omega^- \pi^+)$ is used to calculate the p_T -independent branching-fraction ratio. The results are averaged over p_T using the inverse uncorrelated relative uncertainties as weights. It is measured to be 1.12 ± 0.22 (stat.) ± 0.27 (syst.). This ratio is compared with previous measurements from the CLEO and Belle Collaborations, as well as with theory predictions based on the light-front approach, light cone sum rule, and the effective colour approach under the framework of SU(3)_f symmetry calculations. The ratio measured in this thesis is compatible within 1σ with the CLEO result and it is 2.3σ lower than the Belle result. Once there is a measurement available for the absolute branching fraction $BR(\Omega_c^0 \rightarrow \Omega^- \pi^+)$, the result presented in this thesis will allow the extraction of the absolute value of $BR(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e)$ and thereby the production cross section of Ω_c^0 baryons in semileptonic decay channel. Moreover, the contribution to total $c\bar{c}$ production cross section at $\sqrt{s} = 13$ TeV and fragmentation fraction of Ω_c^0 can be also obtained.

Currently, we do not have yet a thorough understanding of the hadronisation process of heavy-flavour quarks at high energies. The two analyses reported in this thesis offer valuable insights into charm-strange baryon hadronisation mechanisms, exploring the role of strange quarks or strange diquarks. These analyses also contribute to ongoing investigations into the description of charm-strange baryon production in the colour reconnection approach and the possible contribution of coalescence to charm quark hadronisation in pp collisions. Notably, the Ξ_c^0/D^0 and $BR \times \Omega_c^0/D^0$ baryon-to-meson ratios are best described by the Catania model, where the contribution from the fragmentation process is sub-leading. This implies that pure coalescence mechanisms play a major role in the strange-charm baryon hadronisation process. Moreover, these measurements can provide further constraints on theoretical calculations, and serve as fundamental references for investigating the interaction of the charm and beauty quark with the medium formed in heavy-ion collisions.

In both of the analyses presented in this thesis, discrepancies between experimental measurements and



theoretical predictions have been observed. Besides that, the recent discovery of enhanced baryon production in hadronic collision systems, as well as other observed features similar to those in Pb–Pb collisions, such as collective behaviour and strangeness enhancement, raises questions about the conventional view that a QGP-like medium is only formed in nucleus-nucleus systems. Therefore, further experimental and theoretical work is needed to fully understand the charm quark hadronisation mechanisms and the underlying dynamics in hadronic collisions at the LHC in the future. It may be addressed by inspecting with the upgraded ALICE experiments (ALICE 2 and ALICE 3).

5.2 Outlook

There are already many performance study results achieved for ALICE 2 and projections for ALICE 3. In the following, selected results of baryons are presented, further advancing our understanding of heavy flavour physics. In the end, the perspectives of new measurements that could be accessible in the upgrade of the ALICE 2 and ALICE 3 are briefly discussed.

Performance with ALICE 2 and ALICE 3

The upgraded ALICE 2, discussed in Section 2.5, with improved track reconstruction performance, will make comprehensive high-precision measurements of charm baryon production yield across collision systems down to low p_T in Run 3 and 4.

The first new results of charm baryons Λ_c^+ and Ξ_c^0 from Run 3 data in pp collisions at $\sqrt{s} = 13.6$ TeV are shown in Fig. 5.1. It is expected to carry out the measurements over a broad momentum range, which will provide new insights into the understanding of hadronisation mechanisms. In addition, a simulation study of beauty hadron is obtained as well, seen in the bottom panel of Fig. 5.1, which gives us information on the hadronisation effects in the beauty sector.

The ALICE 3 experiment, introduced in Section 2.5.3, is specially designed to track weakly decaying particles prior to their decays. A new method, called ‘strangeness tracking’, is implemented to improve the reconstruction of weakly decaying particles that come from multi-charm baryon decays (also implemented in ALICE 2), resulting from high-precision silicon pixel tracking layers to increase the tracking efficiency.

$$\Xi_{cc}^{++} \rightarrow \Xi_c^+ + \pi^+ \rightarrow \Xi^- + 3\pi^+ \text{ and } \Omega_{cc}^+ \rightarrow \Omega_c^0 + \pi^+ \rightarrow \Omega^- + 2\pi^+ \quad (5.1)$$

In Fig. 5.2, the evaluated simulation study from PYTHIA 8 of the Ξ_{cc}^{++} and Ω_{cc}^+ baryons with ALICE 3 is reported, which is obtained from central Pb–Pb collisions with a total sampled luminosity of (35 nb^{-1}).

The combined advancements of increased strangeness tracking efficiency and the improved impact pa-

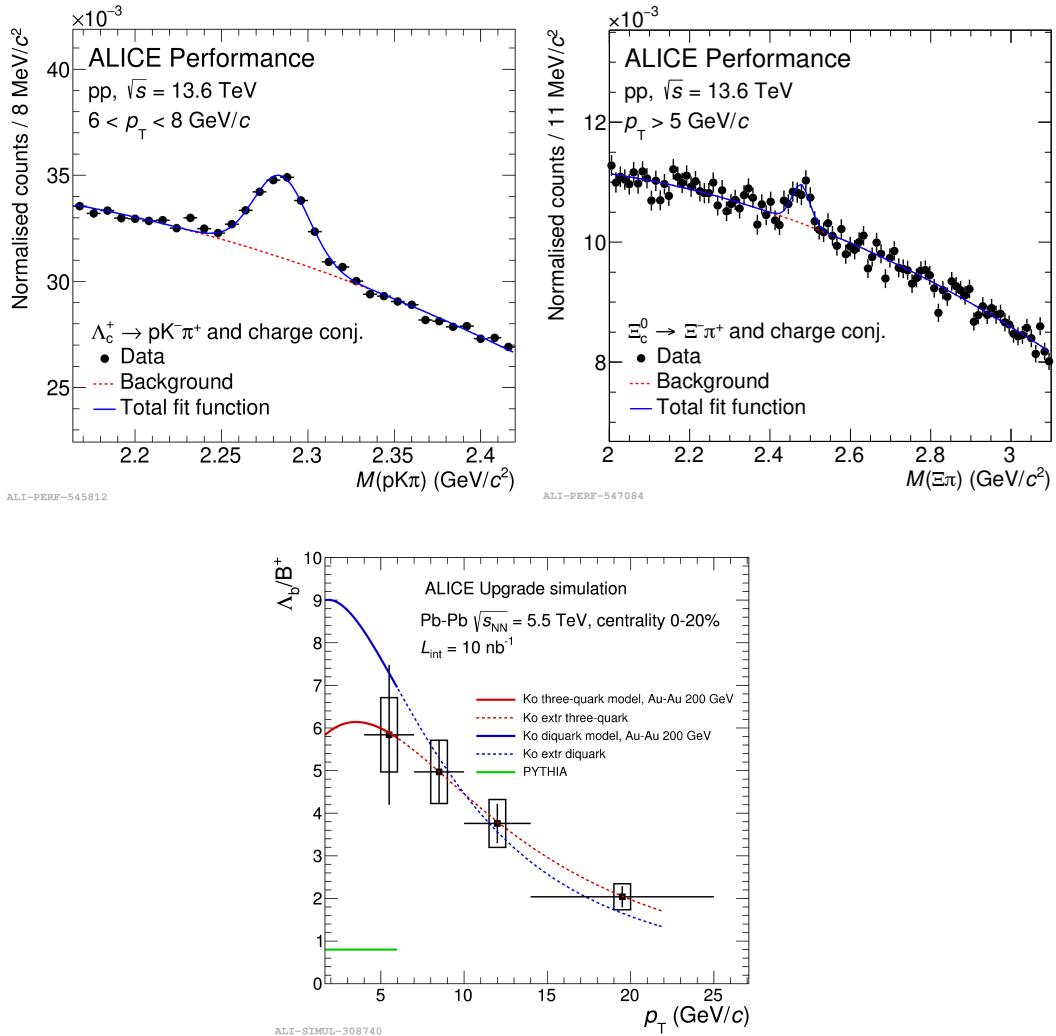


Figure 5.1: Top panels: Invariant-mass distribution of baryons in pp collisions at $\sqrt{s} = 13.6$ TeV. The background distribution is modeled with a second-order polynomial, while the signal with a Gaussian function. Left panel: The invariant mass distribution of baryon Λ_c^+ with $6 < p_T < 8$ GeV/c. Right panel: The invariant mass distribution of Ξ_c^0 candidates from hadronic decay channel $\Xi^- \pi^+$ with $p_T > 5$ GeV/c. Bottom panel: Yield ratio of Λ_b/B^+ as a function of p_T in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.52$ TeV. Simulation for Run 3 + Run 4. Figures from [43].

parameter resolution in ALICE 3 will lead to a much improved capability of heavy-flavour hadron reconstruction.

Multi-charm hadron states

Measurements of multi-charm hadrons, such as Ξ_{cc}^+ ($cc\bar{d}$), Ξ_{cc}^{++} ($cc\bar{u}$), Ω_{cc}^+ ($cc\bar{s}$), Ω_{ccc}^{++} (ccc), and recently discovered exotic states like T_{cc}^+ ($cc\bar{u}\bar{d}$), would offer direct insights into hadron formation from a deconfined quark-gluon plasma. Due to recombination effects, the yields of multi-charm baryons relative to the number of produced charm quarks are predicted to be significantly enhanced in heavy-ion

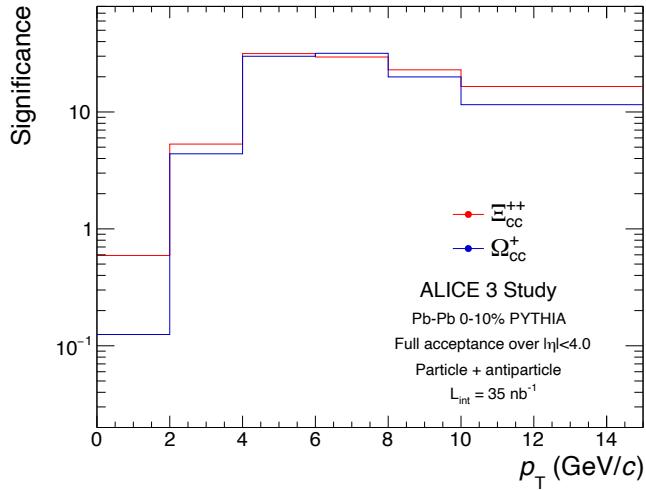


Figure 5.2: Expected Ξ_{cc}^{++} and Ω_{cc}^+ significance in 0-10% central Pb–Pb collisions at $\sqrt{s_{NN}} = 5.52$ TeV as a function of p_T with a 0.2 T magnetic field. Figures from [43].

collisions compared to the ones in pp collisions. Most recently, the SHMc and the Catania models provided predictions for the production of multi-charm hadrons in Pb–Pb collisions, exploring the possible system size dependence via Kr-Kr, to Ar-Ar and O-O collisions. Future experimental data, alongside theoretical calcualtions, will enable precise quantification of these effects and test different hadronization mechanisms.

Beauty hadrons

Beauty quarks serve as even better probes of pQCD than charm quarks, due to their larger mass. They offer better constraints on the propagation of heavy quarks in the later stages of the medium evolution. The B_c^+ meson contains a b and c quarks, which is the only hadron observed so far with heavy quarks of two different flavours, making it a unique probe for investigating the effects of hot and dense nuclear matter created in heavy-ion collisions. Furthermore, B-meson production can provide a reference for studies of beauty hadronisation with Λ_b baryons. Comprehensive measurements of beauty hadron production, including Λ_b , Ξ_b , and Ω_b baryons, are crucial for understanding thermalisation processes in the beauty sector and providing benchmarks for theoretical models.

High mass baryon states

Exploring the measurements of higher-mass charm- and beauty-baryon states, not yet reported in PDG lists, is greatly significant for advancing our understanding of hadronisation processes. Recently, there have been new results of $D_{s1}^+(2536)$ and $D_{s1}^{*+}(2573)$ resonances in pp collisions, offer new pieces of



information to those of ground-state charm mesons to investigate the charm-quark hadronisation. Further studies into resonant baryon states will allow the validation of the number of additional baryon resonance states predicted by the relativistic quark model, helping to understand the description of the baryon-to-meson yield ratios in pp as well as heavy-ion collisions.

Rapidity study

Possible rapidity dependence of both the Λ_c^+ and Ξ_c^+ baryon-to-meson ratios are observed by ALICE and LHCb Collaborations in pp, p–Pb and Pb–Pb collisions. ALICE and LHCb Collaborations probe different regions of the kinematic phase space, making a direct comparison and understanding of the differences in the results more complex. In the future, measurements will aim for consistency in centrality and p_T intervals, potentially including heavier charm baryons. Besides that, extending measurements to other rapidity regions in ALICE, such as forward or backward, may provide more comprehensive knowledge of the kinematic distribution of charm-baryon formation.

Furthermore, it would be also interesting to investigate models including coalescence as a function of rapidity, to study if any rapidity dependence of the baryon-to-meson yield ratios would be observed, and to understand if a possible effect would be driven by the medium density and/or by possible differences in the charm density between different rapidity ranges (mid, forward, backward).



References

- [1] P. A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.
- [2] Michelangelo L. Mangano. Two lectures on heavy quark production in hadronic collisions. *Proc. Int. Sch. Phys. Fermi*, 137:95–137, 1998.
- [3] Matteo Cacciari, Mario Greco, and Paolo Nason. The p_T spectrum in heavy-flavour hadroproduction. *JHEP*, 05:007, 1998.
- [4] Matteo Cacciari, Stefano Frixione, and Paolo Nason. The p_T spectrum in heavy flavor photoproduction. *JHEP*, 03:006, 2001.
- [5] B. A. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger. Collinear subtractions in hadroproduction of heavy quarks. *Eur. Phys. J. C*, 41:199–212, 2005.
- [6] B. A. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger. Inclusive D^{*+} -production in p anti-p collisions with massive charm quarks. *Phys. Rev. D*, 71:014018, 2005.
- [7] Rafał Maciąła and Antoni Szczurek. Production of Λ_c baryons at the LHC within the k_T -factorization approach and independent parton fragmentation picture. *Phys. Rev. D*, 98(1):014016, 2018.
- [8] Shreyasi Acharya et al. Measurement of D^0 , D^+ , D^{*+} and D_s^+ production in pp collisions at $\sqrt{s} = 5.02$ TeV with ALICE. *Eur. Phys. J. C*, 79(5):388, 2019.
- [9] R Aaij et al. Measurement of B meson production cross-sections in proton-proton collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 08:117, 2013.
- [10] Shreyasi Acharya et al. Λ_c^+ production in pp and in p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Rev. C*, 104(5):054905, 2021.
- [11] Shreyasi Acharya et al. Λ_c^+ Production and Baryon-to-Meson Ratios in pp and p–Pb Collisions at $\sqrt{s_{NN}} = 5.02$ TeV at the LHC. *Phys. Rev. Lett.*, 127(20):202301, 2021.
- [12] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction. *JHEP*, 09:126, 2007.
- [13] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.



- [14] Mykhailo Lisovyi, Andrii Verbytskyi, and Oleksandr Zenaiev. Combined analysis of charm-quark fragmentation-fraction measurements. *Eur. Phys. J. C*, 76(7):397, 2016.
- [15] R Aaij et al. Prompt charm production in pp collisions at $\sqrt{s} = 7$ TeV. *Nucl. Phys. B*, 871:1–20, 2013.
- [16] Shreyasi Acharya et al. Measurement of beauty and charm production in pp collisions at $\sqrt{s} = 5.02$ TeV via non-prompt and prompt D mesons. *JHEP*, 05:220, 2021.
- [17] Torbjørn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.
- [18] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.
- [19] Shreyasi Acharya et al. First measurement of Λ_c^+ production down to $p_T = 0$ in pp and p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Rev. C*, 107(6):064901, 2023.
- [20] Shreyasi Acharya et al. Measurement of Prompt D^0 , Λ_c^+ , and $\Sigma_c^{0,++}(2455)$ Production in Proton–Proton Collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 128(1):012001, 2022.
- [21] Shreyasi Acharya et al. Measurement of the Cross Sections of Ξ_c^0 and Ξ_c^+ Baryons and of the Branching-Fraction Ratio $\text{BR}(\Xi_c^0 \rightarrow \Xi^- e^+ \nu_e)/\text{BR}(\Xi_c^0 \rightarrow \Xi^- \pi^+)$ in pp collisions at 13 TeV. *Phys. Rev. Lett.*, 127(27):272001, 2021.
- [22] Shreyasi Acharya et al. Charm production and fragmentation fractions at midrapidity in pp collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 12:086, 2023.
- [23] Jesper R. Christiansen and Peter Z. Skands. String Formation Beyond Leading Colour. *JHEP*, 08:003, 2015.
- [24] Vincenzo Minissale, Salvatore Plumari, and Vincenzo Greco. Charm hadrons in pp collisions at LHC energy within a coalescence plus fragmentation approach. *Phys. Lett. B*, 821:136622, 2021.
- [25] Shreyasi Acharya et al. First measurement of Ω_c^0 production in pp collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 846:137625, 2023.
- [26] Esma Mobs. The CERN accelerator complex in 2019. Complexe des accélérateurs du CERN en 2019. 2019. General Photo.



- [27] K Aamodt et al. Alignment of the ALICE Inner Tracking System with cosmic-ray tracks. *JINST*, 5:P03003, 2010.
- [28] J. Alme et al. The ALICE TPC, a large 3-dimensional tracking device with fast readout for ultra-high multiplicity events. *Nucl. Instrum. Meth. A*, 622:316–367, 2010.
- [29] Betty Bezverkhny Abelev et al. Performance of the ALICE Experiment at the CERN LHC. *Int. J. Mod. Phys. A*, 29:1430044, 2014.
- [30] B Abelev et al. Technical Design Report for the Upgrade of the ALICE Inner Tracking System. *J. Phys. G*, 41:087002, 2014.
- [31] ALICE upgrades during the LHC Long Shutdown 2. 2 2023.
- [32] Luciano Musa. Letter of Intent for an ALICE ITS Upgrade in LS3. Technical report, CERN, Geneva, 2019.
- [33] Shreyasi Acharya et al. Measurement of the production cross section of prompt Ξ_c^0 baryons at midrapidity in pp collisions at $\sqrt{s} = 5.02$ TeV. *JHEP*, 10:159, 2021.
- [34] S. Acharya et al. First measurement of Ξ_c^0 production in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett. B*, 781:8–19, 2018.
- [35] Peter Skands, Stefano Carrazza, and Juan Rojo. Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur. Phys. J. C*, 74(8):3024, 2014.
- [36] Jun Song, Hai-hong Li, and Feng-lan Shao. New feature of low p_T charm quark hadronization in pp collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J. C*, 78(4):344, 2018.
- [37] Min He and Ralf Rapp. Charm-Baryon Production in Proton-Proton Collisions. *Phys. Lett. B*, 795:117–121, 2019.
- [38] R. Ammar et al. Observation of the decay $\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e$. *Phys. Rev. Lett.*, 89:171803, 2002.
- [39] Y. B. Li et al. First test of lepton flavor universality in the charmed baryon decays $\Omega_c^0 \rightarrow \Omega^- l^+ \nu_l$ using data of the Belle experiment. *Phys. Rev. D*, 105(9):L091101, 2022.
- [40] Yu-Kuo Hsiao, Ling Yang, Chong-Chung Lih, and Shang-Yuu Tsai. Charmed Ω_c weak decays into Ω in the light-front quark model. *Eur. Phys. J. C*, 80(11):1066, 2020.
- [41] T. M. Aliev, S. Bilmis, and M. Savci. Charmed baryon $\Omega_c^0 \rightarrow \Omega^- l^+ \nu_l$ and $\Omega_c^0 \rightarrow \Omega_c^- \pi^+(\rho^+)$ decays in light cone sum rules. *Phys. Rev. D*, 106(7):074022, 2022.



- [42] Chia-Wei Liu. Nonleptonic two-body weak decays of charmed baryons. *Phys. Rev. D*, 109(3):033004, 2024.
- [43] Letter of intent for ALICE 3: A next-generation heavy-ion experiment at the LHC. 11 2022.
- [44] ALICE DPG. [Analysis Object Tools - Systematic uncertainties on TPC-ITS matching in various Run-2 data samples](#). *ALICE Twiki*, 2019.
- [45] R. L. Workman et al. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.
- [46] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, 1964.
- [47] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, 1964.
- [48] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.
- [49] Serguei Chatrchyan et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012.
- [50] Georges 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*, 716:1–29, 2012.
- [51] H. L. Anderson, E. Fermi, E. A. Long, and D. E. Nagle. Total Cross-sections of Positive Pions in Hydrogen. *Phys. Rev.*, 85:936, 1952.
- [52] David J. Gross and Frank Wilczek. Ultraviolet Behavior of Nonabelian Gauge Theories. *Phys. Rev. Lett.*, 30:1343–1346, 1973.
- [53] H. David Politzer. Reliable Perturbative Results for Strong Interactions? *Phys. Rev. Lett.*, 30:1346–1349, 1973.
- [54] Kenneth G. Wilson. Confinement of Quarks. *Phys. Rev. D*, 10:2445–2459, 1974.
- [55] H. Suganuma, S. Sasaki, and H. Toki. Color confinement, quark pair creation and dynamical chiral symmetry breaking in the dual Ginzburg-Landau theory. *Nucl. Phys. B*, 435:207–240, 1995.
- [56] Yoichiro Nambu and G. Jona-Lasinio. Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. 1. *Phys. Rev.*, 122:345–358, 1961.
- [57] Ta-Pei [0000-0002-1137-0969] Cheng and Ling-Fong [0000-0002-8035-3329] Li. *Gauge Theory of Elementary Particle Physics*. Oxford University Press, Oxford, UK, 1984.



- [58] Yi-Bo Yang, Jian Liang, Yu-Jiang Bi, Ying Chen, Terrence Draper, Keh-Fei Liu, and Zhaofeng Liu. Proton Mass Decomposition from the QCD Energy Momentum Tensor. *Phys. Rev. Lett.*, 121(21):212001, 2018.
- [59] John C. Collins, Davison E. Soper, and George F. Sterman. Factorization of Hard Processes in QCD. *Adv. Ser. Direct. High Energy Phys.*, 5:1–91, 1989.
- [60] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, Pavel M. Nadolsky, and W. K. Tung. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 07:012, 2002.
- [61] Sayipjamal Dulat, Tie-Jiun Hou, Jun Gao, Marco Guzzi, Joey Huston, Pavel Nadolsky, Jon Pumplin, Carl Schmidt, Daniel Stump, and C. P. Yuan. New parton distribution functions from a global analysis of quantum chromodynamics. *Phys. Rev. D*, 93(3):033006, 2016.
- [62] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J. C*, 77(10):663, 2017.
- [63] V. G. Kartvelishvili, A. K. Likhoded, and V. A. Petrov. On the Fragmentation Functions of Heavy Quarks Into Hadrons. *Phys. Lett. B*, 78:615–617, 1978.
- [64] Neelakshi N. K. Borah, D. K. Choudhury, and P. K. Sahariah. Non-singlet spin structure function $g_1(\text{NS})(x,t)$ in the DGLAP approach. *Pramana*, 79:833–837, 2012.
- [65] V. N. Gribov and L. N. Lipatov. e^+e^- pair annihilation and deep inelastic ep scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15:675–684, 1972.
- [66] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys. B*, 126:298–318, 1977.
- [67] S. Catani, M. Ciafaloni, and F. Hautmann. GLUON CONTRIBUTIONS TO SMALL x HEAVY FLAVOR PRODUCTION. *Phys. Lett. B*, 242:97–102, 1990.
- [68] John C. Collins and R. Keith Ellis. Heavy quark production in very high-energy hadron collisions. *Nucl. Phys. B*, 360:3–30, 1991.
- [69] Rafal Maciula and Antoni Szczurek. Open charm production at the LHC - k_t -factorization approach. *Phys. Rev. D*, 87(9):094022, 2013.
- [70] Johannes Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, 76(4):196, 2016.
- [71] Stefano Frixione, Paolo Nason, and Bryan R. Webber. Matching NLO QCD and parton showers in heavy flavor production. *JHEP*, 08:007, 2003.



- [72] M. Klasen, C. Klein-Bösing, K. Kovarik, G. Kramer, M. Topp, and J. Wessels. NLO Monte Carlo predictions for heavy-quark production at the LHC: pp collisions in ALICE. *JHEP*, 08:109, 2014.
- [73] Albert M Sirunyan et al. Production of Λ_c^+ baryons in proton-proton and lead-lead collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Lett. B*, 803:135328, 2020.
- [74] R. Aaij et al. Measurement of b -hadron production fractions in 7 TeV pp collisions. *Phys. Rev. D*, 85:032008, 2012.
- [75] R. Aaij et al. Study of the production of Λ_b^0 and \bar{B}^0 hadrons in pp collisions and first measurement of the $\Lambda_b^0 \rightarrow J/\psi p K^-$ branching fraction. *Chin. Phys. C*, 40(1):011001, 2016.
- [76] Roel Aaij et al. Measurement of b hadron fractions in 13 TeV pp collisions. *Phys. Rev. D*, 100(3):031102, 2019.
- [77] Serguei Chatrchyan et al. Measurement of the Λ_b Cross Section and the $\bar{\Lambda}_b$ to Λ_b Ratio with $J/\Psi\Lambda$ Decays in pp Collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett. B*, 714:136–157, 2012.
- [78] D. Bortoletto et al. Charm Production in Nonresonant e^+e^- Annihilations at $\sqrt{s} = 10.55$ GeV. *Phys. Rev. D*, 37:1719, 1988. [Erratum: *Phys. Rev. D* 39, 1471 (1989)].
- [79] P. Avery et al. Inclusive production of the charmed baryon Λ_c^+ from e^+e^- annihilations $\sqrt{s} = 10.55$ GeV. *Phys. Rev. D*, 43:3599–3610, 1991.
- [80] H. Albrecht et al. Inclusive production of D^0 , D^+ and D^{*+} (2010) mesons in B decays and non-resonant e^+e^- annihilation at $\sqrt{s} = 10$ GeV. *Z. Phys. C*, 52:353–360, 1991.
- [81] H. Albrecht et al. Production of D_s^+ mesons in B decays and determination of $f(D(s))$. *Z. Phys. C*, 54:1–12, 1992.
- [82] H. Albrecht et al. Observation of the Charmed Baryon Λ_c^+ in e^+e^- Annihilation at $\sqrt{s} = 10.6$ GeV. *Phys. Lett. B*, 207:109–114, 1988.
- [83] R. Seuster et al. Charm hadrons from fragmentation and B decays in e^+e^- annihilation at $\sqrt{s} = 10.6$ GeV. *Phys. Rev. D*, 73:032002, 2006.
- [84] Bernard Aubert et al. Measurement of D_s^+ and D_s^{*+} production in B meson decays and from continuum e^+e^- annihilation at $\sqrt{s} = 10.6$ GeV. *Phys. Rev. D*, 65:091104, 2002.
- [85] Bernard Aubert et al. Inclusive Λ_c^+ Production in e^+e^- Annihilations at $\sqrt{s} = 10.54$ GeV and in $\Upsilon(4S)$ Decays. *Phys. Rev. D*, 75:012003, 2007.



- [86] R. Barate et al. Study of charm production in Z decays. *Eur. Phys. J. C*, 16:597–611, 2000.
- [87] Alexander et al. A study of charm hadron production in $Z^0 \rightarrow c\bar{c}$ and $Z^0 \rightarrow b\bar{b}$ decays at LEPdecays at LEP. *Zeitschrift für Physik C: Particles and Fields*, 72:1–16, 1996.
- [88] P. Abreu et al. Measurements of the Z partial decay width into c anti- c and multiplicity of charm quarks per b decay. *Eur. Phys. J. C*, 12:225–241, 2000.
- [89] K. Ackerstaff et al. Measurement of $f(c \rightarrow D^{*+}X)$, $f(b \rightarrow D^{*+}X)$ and $\Gamma_{c\bar{c}}/\Gamma_{\text{had}}$ using $D^{*\pm}$ mesons. *Eur. Phys. J. C*, 1:439–459, 1998.
- [90] P. Abreu et al. Determination of $P(c \rightarrow D^{*+})$ and $BR(c \rightarrow l^+)$ at LEP-1. *Eur. Phys. J. C*, 12:209–224, 2000.
- [91] H. Abramowicz et al. Measurement of charm fragmentation fractions in photoproduction at HERA. *JHEP*, 09:058, 2013.
- [92] S. Chekanov et al. Measurement of charm fragmentation ratios and fractions in photoproduction at HERA. *Eur. Phys. J. C*, 44:351–366, 2005.
- [93] S. Chekanov et al. Measurement of D mesons production in deep inelastic scattering at HERA. *JHEP*, 07:074, 2007.
- [94] H. Abramowicz et al. Measurement of D^+ and Λ_c^+ production in deep inelastic scattering at HERA. *JHEP*, 11:009, 2010.
- [95] A. Aktas et al. Inclusive production of D^0 , D^+ , D_s^+ and D^{*+} mesons in deep inelastic scattering at HERA. *Eur. Phys. J. C*, 38:447–459, 2005.
- [96] Shreyasi Acharya et al. Measurement of D-meson production at mid-rapidity in pp collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J. C*, 77(8):550, 2017.
- [97] Shreyasi Acharya et al. Measurement of prompt D^0 , D^+ , D^{*+} , and D_s^+ production in p-Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV. *JHEP*, 12:092, 2019.
- [98] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton Fragmentation and String Dynamics. *Phys. Rept.*, 97:31–145, 1983.
- [99] Shreyasi Acharya et al. Charm-quark fragmentation fractions and production cross section at midrapidity in pp collisions at the LHC. *Phys. Rev. D*, 105(1):L011103, 2022.



- [100] M. Niiyama et al. Production cross sections of hyperons and charmed baryons from e^+e^- annihilation near $\sqrt{s} = 10.52$ GeV. *Phys. Rev. D*, 97(7):072005, 2018.
- [101] Torbjorn Sjostrand and Maria van Zijl. A Multiple Interaction Model for the Event Structure in Hadron Collisions. *Phys. Rev. D*, 36:2019, 1987.
- [102] Stefan Gieseke, Christian Rohr, and Andrzej Siódmiok. Colour reconnections in Herwig++. *Eur. Phys. J. C*, 72:2225, 2012.
- [103] Vardan Khachatryan et al. Charged Particle Multiplicities in $p\bar{p}$ Interactions at $\sqrt{s} = 0.9, 2.36$, and 7 TeV. *JHEP*, 01:079, 2011.
- [104] Betty Bezverkhny Abelev et al. Multiplicity dependence of the average transverse momentum in $p\bar{p}$, p-Pb, and Pb-Pb collisions at the LHC. *Phys. Lett. B*, 727:371–380, 2013.
- [105] G. Aad et al. Charged-particle multiplicities in $p\bar{p}$ interactions measured with the ATLAS detector at the LHC. *New J. Phys.*, 13:053033, 2011.
- [106] Yongseok Oh, Che Ming Ko, Su Houng Lee, and Shigehiro Yasui. Heavy baryon/meson ratios in relativistic heavy ion collisions. *Phys. Rev. C*, 79:044905, 2009.
- [107] Jaroslav Adam et al. First measurement of Λ_c baryon production in Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV. *Phys. Rev. Lett.*, 124(17):172301, 2020.
- [108] Shreyasi Acharya et al. Constraining hadronization mechanisms with $\Lambda c/D0$ production ratios in Pb–Pb collisions at $s_{NN}=5.02$ TeV. *Phys. Lett. B*, 839:137796, 2023.
- [109] Jaroslav Adam et al. Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions. *Nature Phys.*, 13:535–539, 2017.
- [110] Vardan Khachatryan et al. Measurement of long-range near-side two-particle angular correlations in $p\bar{p}$ collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 116(17):172302, 2016.
- [111] Ryan D. Weller and Paul Romatschke. One fluid to rule them all: viscous hydrodynamic description of event-by-event central $p\bar{p}$, p–Pb and Pb–Pb collisions at $\sqrt{s} = 5.02$ TeV. *Phys. Lett. B*, 774:351–356, 2017.
- [112] Chun Shen, Jean-François Paquet, Gabriel S. Denicol, Sangyong Jeon, and Charles Gale. Collectivity and electromagnetic radiation in small systems. *Phys. Rev. C*, 95(1):014906, 2017.



- [113] Moritz Greif, Carsten Greiner, Björn Schenke, Sören Schlichting, and Zhe Xu. Importance of initial and final state effects for azimuthal correlations in p+Pb collisions. *Phys. Rev. D*, 96(9):091504, 2017.
- [114] C. Peterson, D. Schlatter, I. Schmitt, and Peter M. Zerwas. Scaling Violations in Inclusive e^+e^- Annihilation Spectra. *Phys. Rev. D*, 27:105, 1983.
- [115] Anton Andronic, Peter Braun-Munzinger, Markus K. Köhler, Aleksas Mazeliauskas, Krzysztof Redlich, Johanna Stachel, and Vytautas Vislavicius. The multiple-charm hierarchy in the statistical hadronization model. *JHEP*, 07:035, 2021.
- [116] Shreyasi Acharya et al. Λ_c^+ production in pp collisions at $\sqrt{s} = 7$ TeV and in p-Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV. *JHEP*, 04:108, 2018.
- [117] D. Ebert, R. N. Faustov, and V. O. Galkin. Spectroscopy and Regge trajectories of heavy baryons in the relativistic quark-diquark picture. *Phys. Rev. D*, 84:014025, 2011.
- [118] A. Bazavov et al. The melting and abundance of open charm hadrons. *Phys. Lett. B*, 737:210–215, 2014.
- [119] Thomas Gutsche, Mikhail A. Ivanov, Jürgen G. Körner, and Valery E. Lyubovitskij. Nonleptonic two-body decays of single heavy baryons Λ_Q , Ξ_Q , and Ω_Q ($Q = b, c$) induced by W emission in the covariant confined quark model. *Phys. Rev. D*, 98(7):074011, 2018.
- [120] Hai-Yang Cheng. Nonleptonic weak decays of bottom baryons. *Phys. Rev. D*, 56:2799–2811, 1997. [Erratum: Phys. Rev. D 99, 079901 (2019)].
- [121] Shiyong Hu, Guanbao Meng, and Fanrong Xu. Hadronic weak decays of the charmed baryon Ω_c . *Phys. Rev. D*, 101(9):094033, 2020.
- [122] E. Solovieva et al. Study of Ω_c^0 and Ω_c^{*0} Baryons at Belle. *Phys. Lett. B*, 672:1–5, 2009.
- [123] Kai-Lei Wang, Qi-Fang Lü, Ju-Jun Xie, and Xian-Hui Zhong. Toward discovering the excited Ω baryons through nonleptonic weak decays of Ω_c . *Phys. Rev. D*, 107(3):034015, 2023.
- [124] Chao-Qiang Geng, Chia-Wei Liu, Tien-Hsueh Tsai, and Shu-Wei Yeh. Semileptonic decays of anti-triplet charmed baryons. *Phys. Lett. B*, 792:214–218, 2019.
- [125] C. Q. Geng, Y. K. Hsiao, Chia-Wei Liu, and Tien-Hsueh Tsai. Three-body charmed baryon Decays with SU(3) flavor symmetry. *Phys. Rev. D*, 99(7):073003, 2019.



- [126] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.
- [127] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.
- [128] A. Augusto Alves, Jr. et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.
- [129] LHC Machine. *JINST*, 3:S08001, 2008.
- [130] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.
- [131] P Cortese et al. ALICE: Physics performance report, volume I. *J. Phys. G*, 30:1517–1763, 2004.
- [132] Christian Wolfgang Fabjan et al. ALICE: Physics Performance Report. *J. Phys. G*, 32:1295–2040, 2006.
- [133] Walter Blum, Luigi Rolandi, and Werner Riegler. *Particle detection with drift chambers*. Particle Acceleration and Detection. 2008.
- [134] Ivan Kisiel, Igor Kulakov, and Maksym Zyzak. Standalone first level event selection package for the CBM Experiment. *IEEE Transactions on Nuclear Science*, 60(5):3703–3708, 2013.
- [135] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth. A*, 389:81–86, 1997.
- [136] René Brun, F Bruyant, Federico Carminati, Simone Giani, M Maire, A McPherson, G Patrick, and L Urban. *GEANT: Detector Description and Simulation Tool; Oct 1994*. CERN Program Library. CERN, Geneva, 1993. Long Writeup W5013.
- [137] S. Agostinelli et al. GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.
- [138] A Ferrari, Paola R Sala, A Fassò, and Johannes Ranft. *FLUKA: A multi-particle transport code (program version 2005)*. CERN Yellow Reports: Monographs. CERN, Geneva, 2005.
- [139] P Buncic, M Krzewicki, and P Vande Vyvre. Technical Design Report for the Upgrade of the Online-Offline Computing System. Technical report, 2015.
- [140] B Abelev et al. Upgrade of the ALICE Experiment: Letter Of Intent. *J. Phys. G*, 41:087001, 2014.
- [141] Upgrade of the ALICE Time Projection Chamber. Technical report, 2013.



- [142] P Antonioli, A Kluge, and W Riegler. Technical Design Report for the Muon Forward Tracker. Technical report, 2015.
- [143] P Antonioli, A Kluge, and W Riegler. Upgrade of the ALICE Readout and Trigger System. Technical report, 2013. Presently we require a LHCC-TDR reference number.a later stage we will fill the required information.
- [144] Letter of Intent: A Forward Calorimeter (FoCal) in the ALICE experiment. Technical report, CERN, Geneva, 2020.
- [145] E. Abbas et al. Performance of the ALICE VZERO system. *JINST*, 8:P10016, 2013.
- [146] ALICE 2017 luminosity determination for pp collisions at $\sqrt{s} = 5$ TeV. 2018.
- [147] Shreyasi Acharya et al. Measurements of low- p_T electrons from semileptonic heavy-flavour hadron decays at mid-rapidity in pp and Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *JHEP*, 10:061, 2018.
- [148] Shreyasi Acharya et al. Measurement of electrons from semileptonic heavy-flavour hadron decays at midrapidity in pp and Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Lett. B*, 804:135377, 2020.
- [149] Shreyasi Acharya et al. Centrality and transverse momentum dependence of inclusive J/ψ production at midrapidity in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Lett. B*, 805:135434, 2020.
- [150] Shreyasi Acharya et al. Inclusive J/ψ production at mid-rapidity in pp collisions at $\sqrt{s} = 5.02$ TeV. *JHEP*, 10:084, 2019.
- [151] Shreyasi Acharya et al. Multiplicity dependence of (multi-)strange hadron production in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 80(2):167, 2020.
- [152] G. D'Agostini. A Multidimensional unfolding method based on Bayes' theorem. *Nucl. Instrum. Meth. A*, 362:487–498, 1995.
- [153] Tim Adye. Unfolding algorithms and tests using RooUnfold. In *PHYSTAT 2011*, pages 313–318, Geneva, 2011. CERN.
- [154] Constantino Tsallis. Possible Generalization of Boltzmann-Gibbs Statistics. *J. Statist. Phys.*, 52:479–487, 1988.
- [155] Matteo Cacciari, Stefano Frixione, Nicolas Houdeau, Michelangelo L. Mangano, Paolo Nason, and Giovanni Ridolfi. Theoretical predictions for charm and bottom production at the LHC. *JHEP*, 10:137, 2012.



- [156] R. Barate et al. Measurement of the B baryon lifetime and branching fractions in Z decays. *Eur. Phys. J. C*, 2:197–211, 1998.
- [157] René Brun, F. Bruyant, Federico Carminati, Simone Giani, M. Maire, A. McPherson, G. Patrick, and L. Urban. GEANT Detector Description and Simulation Tool. 10 1994.
- [158] Roger John Barlow. Practical statistics for particle physics. *CERN Yellow Rep. School Proc.*, 5:149–197, 2020.
- [159] Andreas Hocker and Vakhtang Kartvelishvili. SVD approach to data unfolding. *Nucl. Instrum. Meth. A*, 372:469–481, 1996.
- [160] Roel Aaij et al. Measurement of the mass and production rate of Ξ_b^- baryons. *Phys. Rev. D*, 99(5):052006, 2019.
- [161] Xing-rui Gou, Feng-lan Shao, Rui-qin Wang, Hai-hong Li, and Jun Song. New insights into hadron production mechanism from p_T spectra in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev. D*, 96(9):094010, 2017.
- [162] R. Frühwirth and R. K. Bock. *Data analysis techniques for high-energy physics experiments*, volume 11. Cambridge University Press, 2000.
- [163] S. Gorbunov and I. Kisel. Elastic net for stand-alone RICH ring finding. *Nucl. Instrum. Meth. A*, 559:139–142, 2006.
- [164] I. Kisel. Event reconstruction in the CBM experiment. *Nucl. Instrum. Meth. A*, 566:85–88, 2006.
- [165] Sergey Gorbunov. *On-line reconstruction algorithms for the CBM and ALICE experiments*. PhD thesis, Goethe U., Frankfurt (main), Frankfurt U., 2013.
- [166] Alexander Radovic, Mike Williams, David Rousseau, Michael Kagan, Daniele Bonacorsi, Alexander Himmel, Adam Aurisano, Kazuhiko Terao, and Taritree Wongjirad. Machine learning at the energy and intensity frontiers of particle physics. *Nature*, 560(7716):41–48, 2018.
- [167] Kim Albertsson et al. Machine Learning in High Energy Physics Community White Paper. *J. Phys. Conf. Ser.*, 1085(2):022008, 2018.
- [168] Sergey Gorbunov. *On-line reconstruction algorithms for the CBM and ALICE experiments*. PhD thesis, Universitätsbibliothek Johann Christian Senckenberg, 2013.
- [169] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, 03 1960.



- [170] Ioannis Antoniou, S.V. Gorbunov, Victor Ivanov, I.V. Kisel, and E.V. Konotopskaya. Elastic neural nets for the traveling salesman problem. *Journal of Computational Methods in Sciences and Engineering*, 2:111–115, 02 2002.
- [171] Maksym Zyzak. *Online selection of short-lived particles on many-core computer architectures in the CBM experiment at FAIR*. PhD thesis, Goethe U., Frankfurt (main), Frankfurt U., 2016.
- [172] I.V. Gorbunov, Sergey and Konotopskaya. Reconstruction of decayed particles based on the kalman filter.
- [173] Bruce H. Denby. Neural Networks and Cellular Automata in Experimental High-energy Physics. *Comput. Phys. Commun.*, 49:429–448, 1988.
- [174] Tianqi Chen and Carlos Guestrin. XGBoost. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, aug 2016.
- [175] Luca Barioglio, Fabio Catalano, Matteo Concas, Pietro Fecchio, Fabrizio Grossa, Francesco Mazzaschi, and Maximiliano Puccio. hipe4ml/hipe4ml, November 2021.
- [176] Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, 2001.
- [177] Byron P. Roe, Hai-Jun Yang, Ji Zhu, Yong Liu, Ion Stancu, and Gordon McGregor. Boosted decision trees, an alternative to artificial neural networks. *Nucl. Instrum. Meth. A*, 543(2-3):577–584, 2005.
- [178] Alan S. Cornell, Wesley Doorsamy, Benjamin Fuks, Gerhard Harmsen, and Lara Mason. Boosted decision trees in the era of new physics: a smuon analysis case study. *JHEP*, 04:015, 2022.
- [179] Rafael G. Mantovani, Tomáš Horváth, Ricardo Cerri, Joaquin Vanschoren, and André C.P.L.F. de Carvalho. Hyper-parameter tuning of a decision tree induction algorithm. In *2016 5th Brazilian Conference on Intelligent Systems (BRACIS)*, pages 37–42, 2016.
- [180] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. 3 2016.
- [181] L. Breiman, Jerome H. Friedman, Richard A. Olshen, and C. J. Stone. Classification and regression trees. *Biometrics*, 40:874, 1984.
- [182] Yann Coadou. Boosted Decision Trees and Applications. *EPJ Web Conf.*, 55:02004, 2013.
- [183] Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals Statist.*, 29(5):1189–1232, 2001.



- [184] ALICE 2016-2017-2018 luminosity determination for pp collisions at $\sqrt{s} = 13$ TeV. 2021.
- [185] Roel Aaij et al. Measurement of the Ω_c^0 baryon lifetime. *Phys. Rev. Lett.*, 121(9):092003, 2018.
- [186] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A Next-generation Hyperparameter Optimization Framework. 7 2019.
- [187] Shashank Shekhar, Adesh Bansode, and Asif Salim. A comparative study of hyper-parameter optimization tools, 2022.
- [188] Gareth James et al. An introduction to statistical learning, 2013.
- [189] Nakul Soni and Jignesh Pandya. Semi-leptonic and pionic decays of Doubly Strange baryons. *DAE Symp. Nucl. Phys.*, 60:694–695, 2015.
- [190] Massimiliano Bonamente. *Statistics and Analysis of Scientific Data*. Graduate Texts in Physics. Springer-Verlag New York, 2013.
- [191] Zhen-Xing Zhao. Weak decays of heavy baryons in the light-front approach. *Chin. Phys. C*, 42(9):093101, 2018.
- [192] Qi-An Zhang et al. First lattice QCD calculation of semileptonic decays of charmed-strange baryons Ξ_c . *Chin. Phys. C*, 46(1):011002, 2022.
- [193] R. N. Faustov and V. O. Galkin. Semileptonic Ξ_c baryon decays in the relativistic quark model. *Eur. Phys. J. C*, 79(8):695, 2019.
- [194] Dpg - analysis object tools - systematic uncertainties on tpc-its matching in various run-2 data samples. <https://twiki.cern.ch/twiki/bin/viewauth/ALICE/AliDPGtoolsTrackSystematicUncertaintyBookkeeping>.



Appendix A

A.1 Ξ_c^0 analysis

A.1.1 Systematic: ITS–TPC matching

A two-step procedure is adopted to propagate the systematic uncertainty on electron tracks to Ξ_c^0 .

First, the systematic uncertainty of ITS-TPC matching efficiency of electrons listed in Tab. A.1, is propagated to the $e\Xi$ pairs at each $p_T^{e\Xi}$ bin, as follows the Eq. A.1.

Table A.1: ITS-TPC matching efficiency for electron taken from DPG [44].

p_T bins	2.0 - 3.0	3.0 - 4.0	4.0 - 6.0	6.0 - 8.0
systematic value	0.026	0.0262	0.0244	0.0266

$$\text{Syst}_{\text{ITS-TPC matching}}^{e\Xi} = \frac{\sum_{i=1}^N (\text{Syst}_i^{e\leftarrow\Xi_c^0} * \text{BC}_i)}{\sum_{i=1}^N \text{BC}_i}, \quad (\text{A.1})$$

where $\text{Syst}_i^{e\leftarrow\Xi_c^0}$ is the one from Tab. A.1 and BC_i represents the i^{th} bin content in $p_T^{e\leftarrow\Xi_c^0}$ distribution at each $p_T^{e\Xi}$ bin, as shown in Fig. A.1, $p_T^{e\leftarrow\Xi_c^0}$ distributions in different $p_T^{e\Xi}$ intervals.

The first step of the systematic uncertainty of ITS-TPC matching efficiency propagated from p_T^e to $p_T^{e\Xi}$ is achieved, which is listed in Tab. A.2, and detailed calculations are illustrated in Eqs. A.2.

$$\text{Syst}_{\text{Bin}[2,3]}^{e\Xi} = \frac{575 * 2.03\% + 23 * 2.6\% + 1 * 2.62\%}{575 + 23 + 1} \approx 2.05\%, \quad (\text{A.2})$$

$$\text{Syst}_{\text{Bin}[3,4]}^{e\Xi} = \frac{610 * 2.03\% + 70 * 2.06\% + 5 * 2.62\%}{610 + 70 + 5} \approx 2.04\%,$$

$$\text{Syst}_{\text{Bin}[4,5]}^{e\Xi} = \frac{215 * 2.03\% + 73 * 2.06\% + 14 * 2.62\%}{215 + 73 + 14} \approx 2.06\%,$$

$$\text{Syst}_{\text{Bin}[5,6]}^{e\Xi} = \frac{73 * 2.03\% + 43 * 2.6\% + 16 * 2.62\% + 10 * 2.44\% + 1 * 2.44\%}{73 + 43 + 16 + 10 + 1} \approx 2.14\%,$$

$$\text{Syst}_{\text{Bin}[6,8]}^{e\Xi} = \frac{30 * 2.03\% + 27 * 2.6\% + 6 * 2.62\% + 13 * 2.44\% + 4 * 2.44\% + 1 * 2.66\%}{30 + 27 + 6 + 13 + 4 + 1} \approx 2.36\%,$$

Second, a similar procedure is applied to propagate the systematic of ITS-TPC matching efficiency of $p_T^{e\Xi}$ to $p_T^{\Xi_c^0}$. The $p_T^{\Xi_c^0}$ distributions in different $p_T^{\Xi_c^0}$ intervals, shown in Fig. A.2, is obtained by projecting Y-axis of response matrix seen in the left panel of Fig. 3.13 of Section 3.3.3.

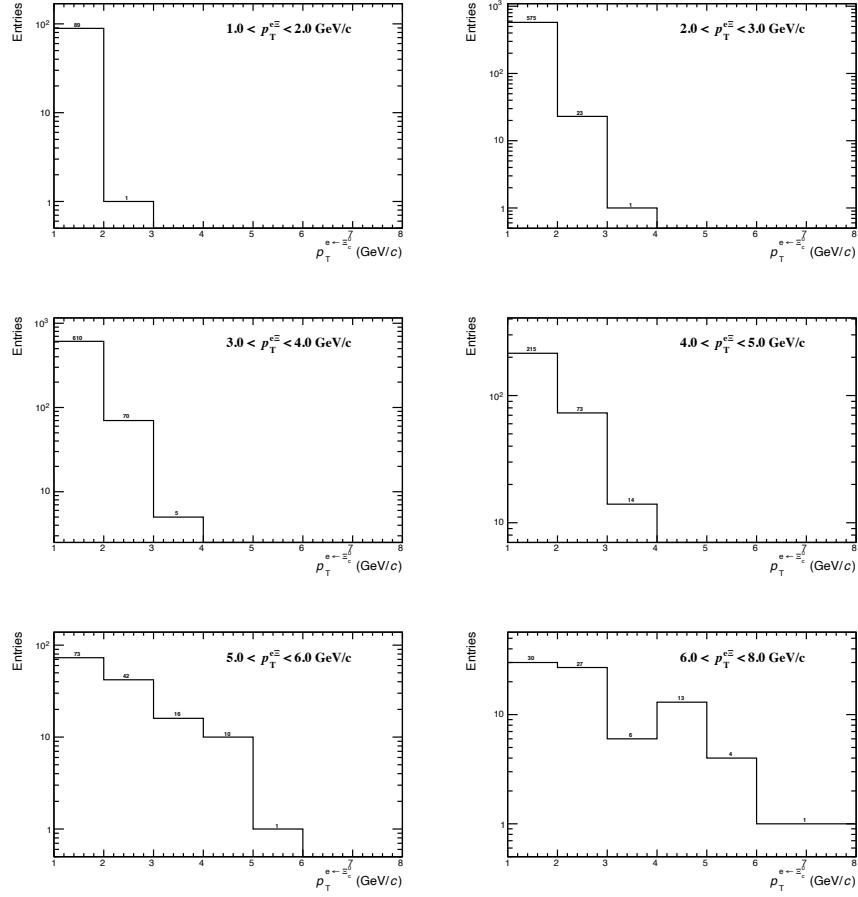


Figure A.1: The $p_T^{e \leftarrow \Xi_c^0}$ distributions in different $p_T^{e \Xi}$ intervals.

Table A.2: The systematic uncertainty of ITS-TPC matching efficiency of $p_T^{e \Xi}$ propagated from p_T^e .

$e\Xi$					
p_T bins (MeV/c)	2-3	3-4	4-5	5-6	6-8
Systematic	2.05%	2.04%	2.06%	2.14%	2.36%

Together with the Eqs. A.4, finally, the systematic uncertainty of ITS-TPC matching efficiency of Ξ_c^0 is assigned as 2% independent of Ξ_c^0 p_T , reported in Tab. A.3.

$$\text{Syst}_{\text{ITS-TPCmatching}}^{\Xi_c^0} = \frac{\sum_{i=1}^N (\text{Syst}_i^{e\Xi \leftarrow \Xi_c^0} * \text{BC}_i)}{\sum_{i=1}^N \text{BC}_i}, \quad (\text{A.3})$$

where $\text{Syst}_i^{e\Xi \leftarrow \Xi_c^0}$ is one from Tab. A.2, and BC_i is the i^{th} bin content in $p_T^{e\Xi}$ distribution at each $p_T^{\Xi_c^0}$ bin as seen in Fig. A.2.

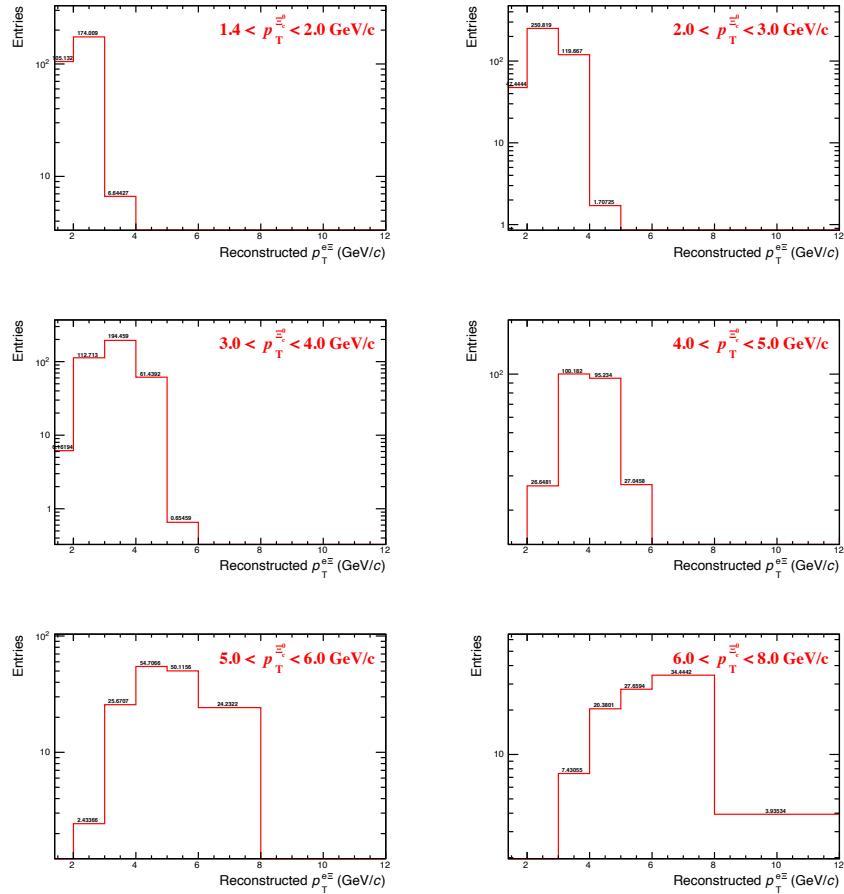


Figure A.2: The $p_T^{e\Xi}$ distributions in different $p_T^{\Xi_c}$ intervals.

$$\text{Syst}_{\text{Bin}[2,3]}^{\Xi_c} = \frac{250.819 * 2.05\% + 119.667 * 2.04\% + 1.70725 * 2.06\%}{250.819 + 119.667 + 1.70725} \approx 2.05\%, \quad (\text{A.4})$$

$$\text{Syst}_{\text{Bin}[3,4]}^{\Xi_c} = \frac{112.7 * 2.05\% + 194.5 * 2.04\% + 61.4 * 2.06\% + 0.65 * 2.14\%}{112.7 + 194.5 + 61.4 + 0.65} \approx 2.05\%,$$

$$\text{Syst}_{\text{Bin}[4,5]}^{\Xi_c} = \frac{26.6 * 2.05\% + 100.2 * 2.04\% + 95.2 * 2.06\% + 27 * 2.14\%}{26.6 + 100.2 + 95.2 + 27} \approx 2.06\%,$$

$$\text{Syst}_{\text{Bin}[5,6]}^{\Xi_c} = \frac{2.4 * 2.05\% + 25.7 * 2.04\% + 54.7 * 2.06\% + 50.1 * 2.14\% + 24.2 * 2.36\%}{2.4 + 25.7 + 54.7 + 50.1 + 24.2} \approx 2.13\%,$$

$$\text{Syst}_{\text{Bin}[6,8]}^{\Xi_c} = \frac{7.4 * 2.04\% + 20.4 * 2.06\% + 27.7 * 2.14\% + 34.4 * 2.36\%}{7.4 + 20.4 + 27.7 + 34.4} \approx 2.2\%,$$

A.1.2 Systematic: Electron identification

Fig. A.3 is the $n\sigma_{\text{TPC}}^{\text{ele}}$ distribution as a function of the electron p_T after applying the particle identification criteria on the TOF signal. Varied criteria were applied on the TPC dE/dx signal to select electron candidates for the systematic checks, which are shown in different colour curves.



Table A.3: The systematic uncertainty of $p_T^{\Xi_c^0}$ propagated from $p_T^{e\Xi}$.

	Ξ_c^0				
p_T bins (MeV/c)	2-3	3-4	4-5	5-6	6-8
Systematic	2.05%	2.05%	2.06%	2.13%	2.2%

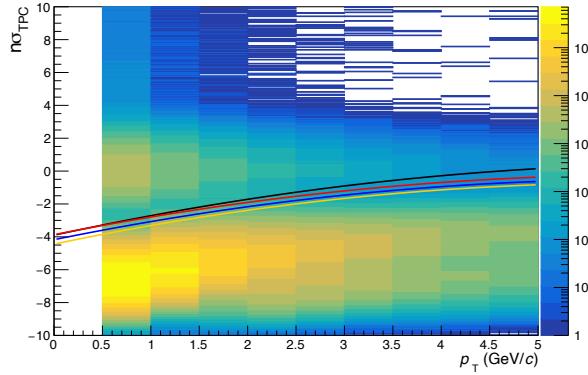


Figure A.3: The varied criteria of the TPC dE/dx signal to select candidates, $n\sigma_{\text{TPC}}^{\text{ele}}$, is applied in this systematic study.

A.2 Ω_c^0 analysis

A.2.1 Traditional rectangular analysis method

Before moving this Ω_c^0 analysis with the implementation of machine learning, the same method as Ξ_c^0 discussed in Section 3, is firstly adopted, to have a look at the feasibility. In the following, the selections and part of the results are briefly presented.

A.2.1.1 Selections for candidates

In this part, the selections for candidates used to reconstruct the Ω_c^0 are listed in Tab A.4, Tab A.5 and Tab A.6.



Table A.4: eID cuts applied in this analysis.

Cuts vairbales	cuts
AOD Filter Bit	4 (Standard cuts with very loose DCA)
Number of CrossedRows	> 70
CrossedRows Over Findable Cluster	> 0.8
Number of TPC PID clusters	> 50
Ratio of findable clusters	> 0.6
ITS/TPC refit	TRUE
Number of ITS cluster	≥ 3
$p_T(\text{GeV}/c)$	> 0.5
$ \eta $	< 0.8
SPD hit	both
$ n\sigma_{\text{TOF}}^{\text{ele}} $	< 3
$ n\sigma_{\text{TPC}}^{\text{ele}} $	$> -3.9 + 1.17 \times x - 0.094 \times x^2$
prefilter cut	$m_{e^+e^-} < 0.05 \text{ GeV}/c^2$

Table A.5: Loose eID cuts applied for the prefilter procedure in this analysis.

Cuts vairbales	cuts
AOD Filter Bit	4
$ n\sigma_{\text{TOF}}^{\text{ele}} $	< 5



Table A.6: The track and topology selections for Ω applied in this analysis.

Cuts variables	cuts
Number of CrossedRows	> 70
CrossedRows Over Findable Cluster	> 0.8
Λ Mass tolerance (MeV/c^2)	7.5
Ω Mass tolerance (MeV/c^2)	6
Competing Mass (to reject Ξ) (MeV/c^2)	8
DCA of V0 to PV (cm)	> 0.088
DCA of V0 daughters to PV (cm)	> 0.073
DCA of bachelor track to PV (cm)	> 0.049
DCA of cascade daughters (cm)	< 1.1033
DCA of V0 daughters (cm)	< 1.116
V0 decay length (cm)	> 0.6
V0 cosine of pointing angle to Ξ vertex	> 0.97
Ω cosine of pointing angle to PV	> 0.985
Ω decay length (cm)	> 0.4
$ n\sigma_{\text{TPC}} $ (proton)	< 4
$ n\sigma_{\text{TPC}} $ (pion)	< 4
$ n\sigma_{\text{TPC}} $ (kaon)	< 4



A.2.1.2 Raw yield extraction and efficiency correction

After applying the selections reported in Tab. A.4 Tab. A.5, and Tab. A.6 in Section A.2.1.1, the invariant-mass distribution of $e\Omega$ RS, and WS pairs can be obtained, as shown in Fig. A.4. To obtain the raw yield, the invariant-mass distribution of $e\Omega$ WS pairs has to be subtracted from the $e\Omega$ RS pairs, seen in Fig. A.5.

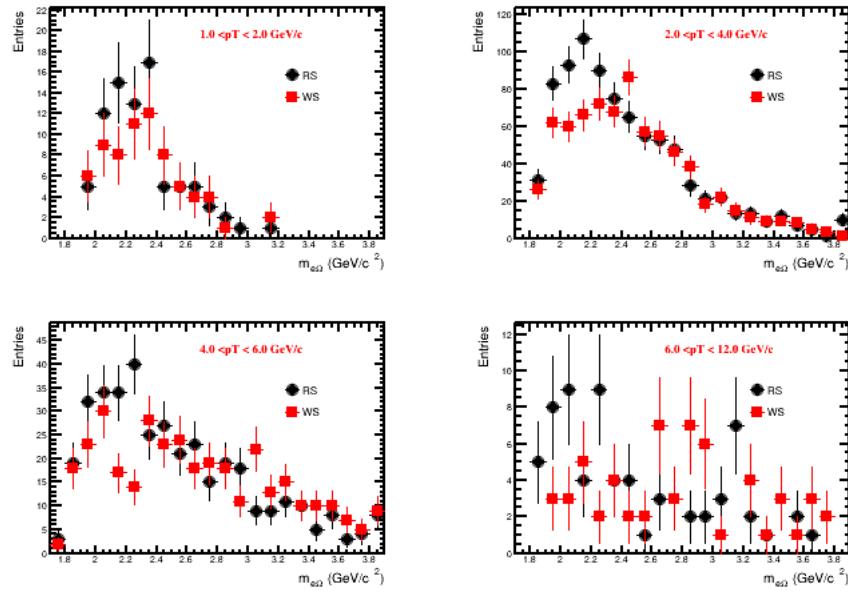


Figure A.4: The invariant mass distribution of $e\Omega$ pairs for RS and WS in SE in each p_T interval.

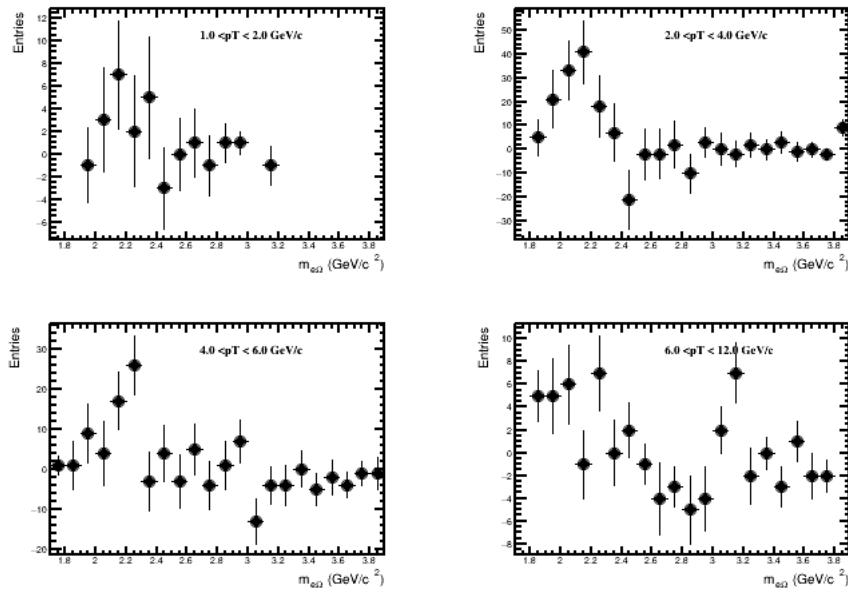


Figure A.5: The invariant mass distribution of $e\Omega$ pairs after the WS mass subtraction in each p_T interval.

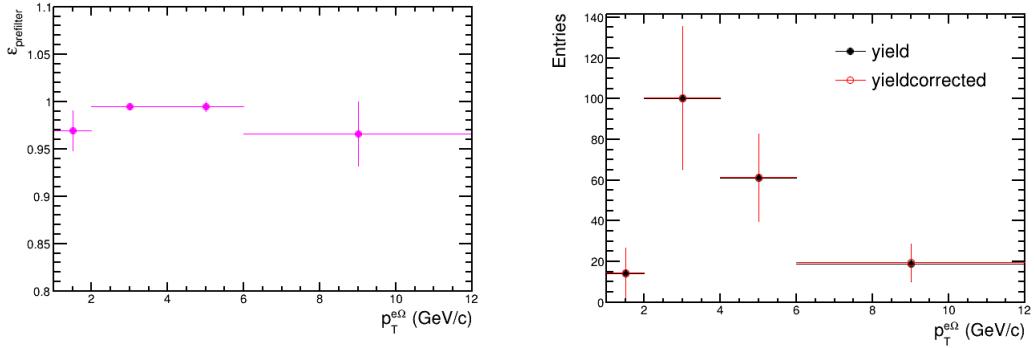


Figure A.6: Left: The prefilter efficiency. Right: The comparison of yield distribution of $p_T^{\text{e}\Omega}$.

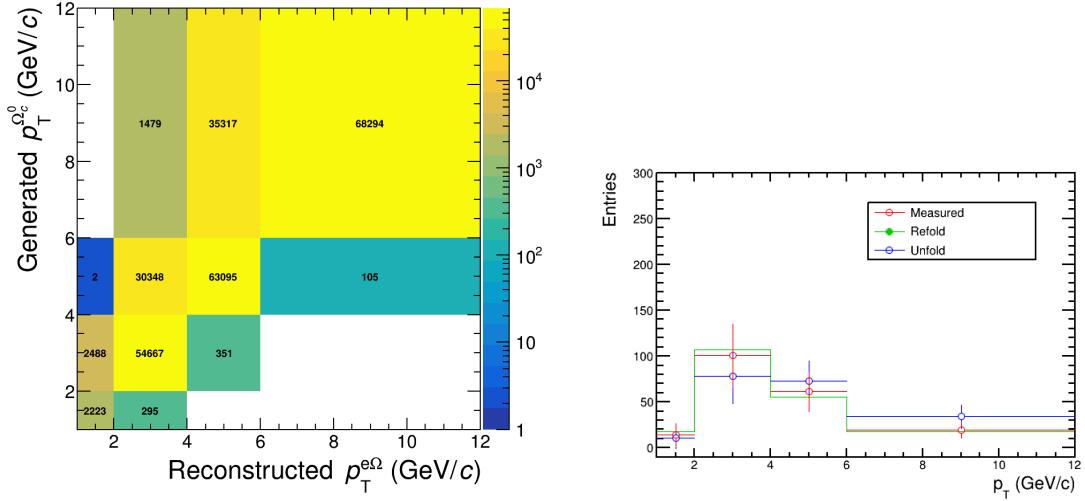


Figure A.7: Left: The correlation matrix between the generated Ω_c^0 -baryon p_T and the reconstructed $e\Omega$ pair p_T . Right: Comparisons of the measured spectrum to the refolded and the unfolded one with and without weights.

As discussed in Section 3.3.1, the prefilter efficiency is used to correct the wrongly tagged electron candidates, illustrated in the left panel of Fig. A.6, and with this, the raw yield comparison is shown in the right panel. The red color is used in the next unfolding procedure.

In Fig. A.7, the left panel shows the response matrix used for the unfolding procedure, and the right panel is the raw yield comparison.

The reconstruction efficiency is shown in the left panel of Fig. A.8, which is used to correct for the unfolded yield, then the p_T -differential production cross sections of inclusive Ω_c^0 baryons multiplied by the branching ratio into $\Omega^- e^+ \nu_e$ is obtained, as shown in the right panel.

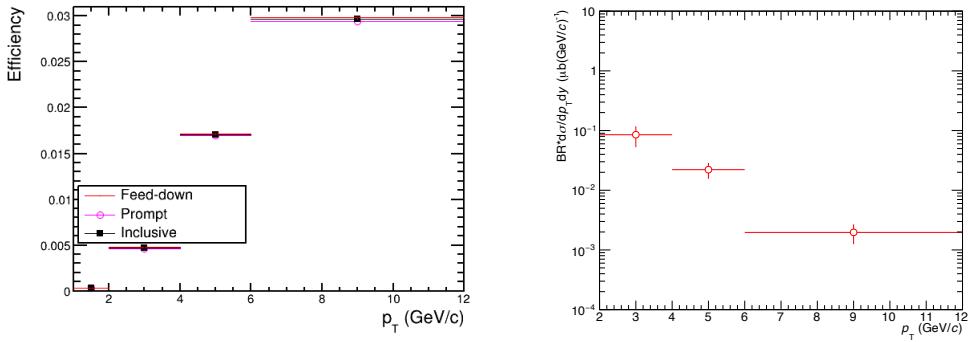


Figure A.8: Left panel: product of acceptance and efficiency for prompt, feed-down, and inclusive Ω_c^0 baryons. Right panel: p_T -differential production cross sections of inclusive Ω_c^0 baryons multiplied by the branching ratio into $\Omega^- e^+ \nu_e$.

A.2.2 Performance with machine learning

At the beginning, the wrong-sign $e\Omega$ pairs in SE are used for the ML training and the background subtraction, to make a straightforward comparison by using the rectangular cuts as shown in Section A.2.1.2. Here the BDT probability output and also the pseudo-significance for the three p_T intervals are shown in Fig. A.9 and Fig. A.10, since the analysed procedure is exactly the same as introduced in Section 4.

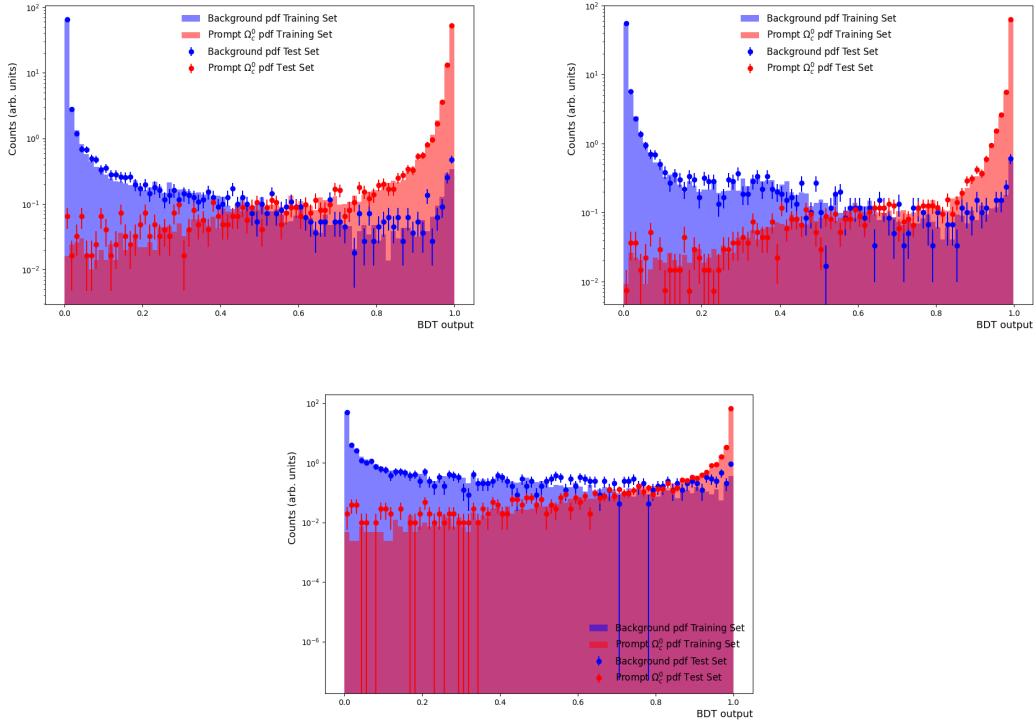


Figure A.9: Machine learning model in each p_T interval in this analysis when SEWS is used for the background. Top left panel: $2 < p_T < 4$ GeV/c. Top right panel: $4 < p_T < 6$ GeV/c. Bottom panel: $6 < p_T < 12$ GeV/c.

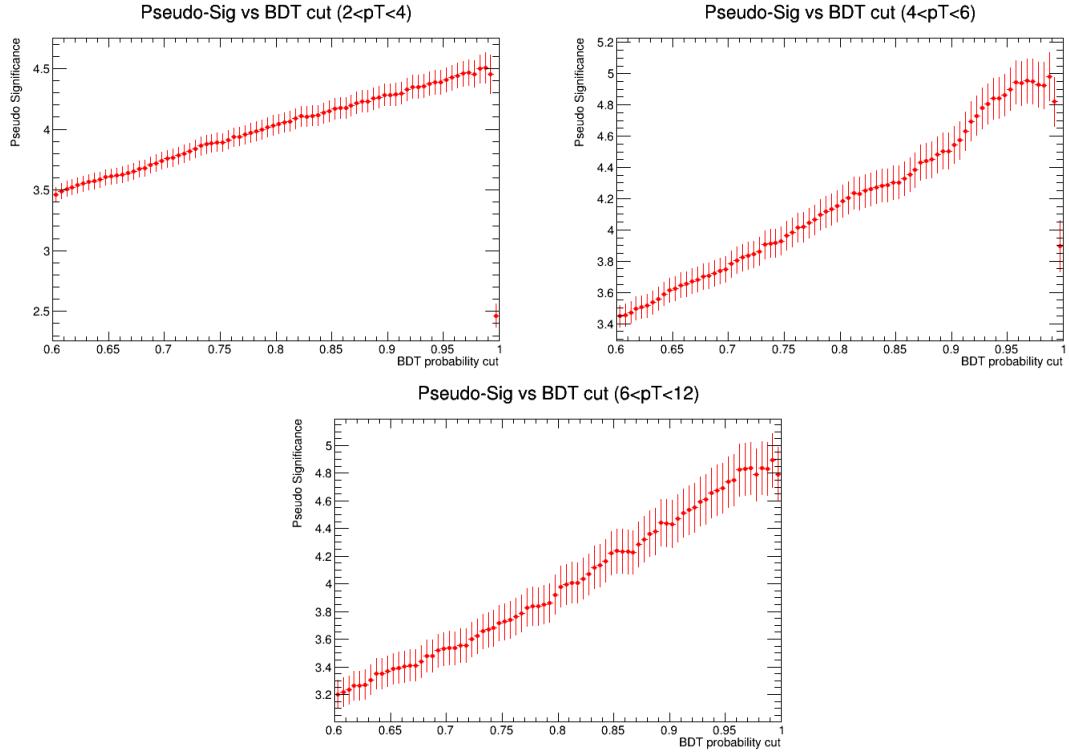


Figure A.10: The pseudo-significance in each p_T interval.

After the application of the BDT selection criteria, the invariant-mass distribution of $e\Omega$ pairs can be obtained as shown in Fig. A.12, and invariant-mass distribution of $e\Omega$ pairs after WS subtracting from RS pairs in each p_T interval is reported in Fig. A.12.

Fig. A.13 shows the response matrix between the generated Ω_c^0 -baryon p_T and the reconstructed $e\Omega$ pair p_T , and also the unfolded raw yield distribution in the right panel.

Fig. A.14 tells the preselection efficiency, BDT efficiency, and the total acceptance times efficiency for the inclusive Ω_c^0 baryons which will be used in the next to correct for the unfolded raw yield.

Fig. A.15 shows the comparison of p_T -differential production cross section of inclusive Ω_c^0 baryons multiplied by the branching ratio into $\Omega^- e^+ \nu_e$, between the traditional rectangular cut method (in red color) and the new ML method (in black color). The improvement of relative statistical uncertainties from ML with respect to the traditional rectangular cut method is illustrated in the right panel, 30% in $2 < p_T < 4$, 20% in $4 < p_T < 6$, and 10% in $6 < p_T < 12$, demonstrating the advantage of using ML.

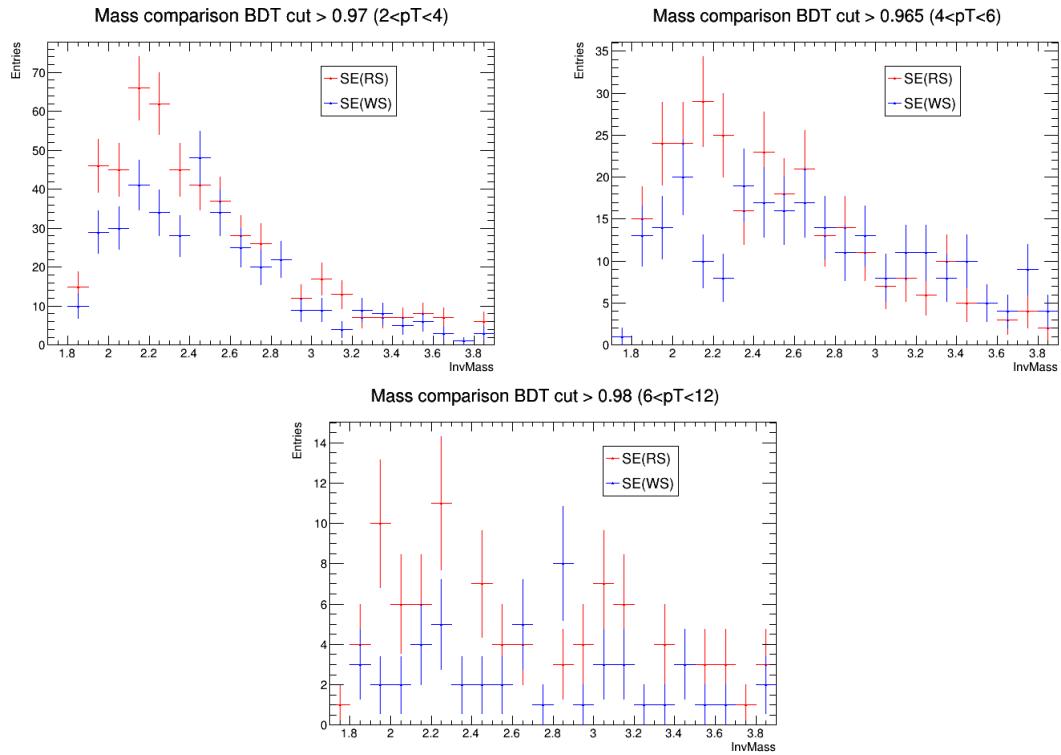


Figure A.11: The invariant-mass distribution of $e\Omega$ RS and WS pairs in each p_T interval.

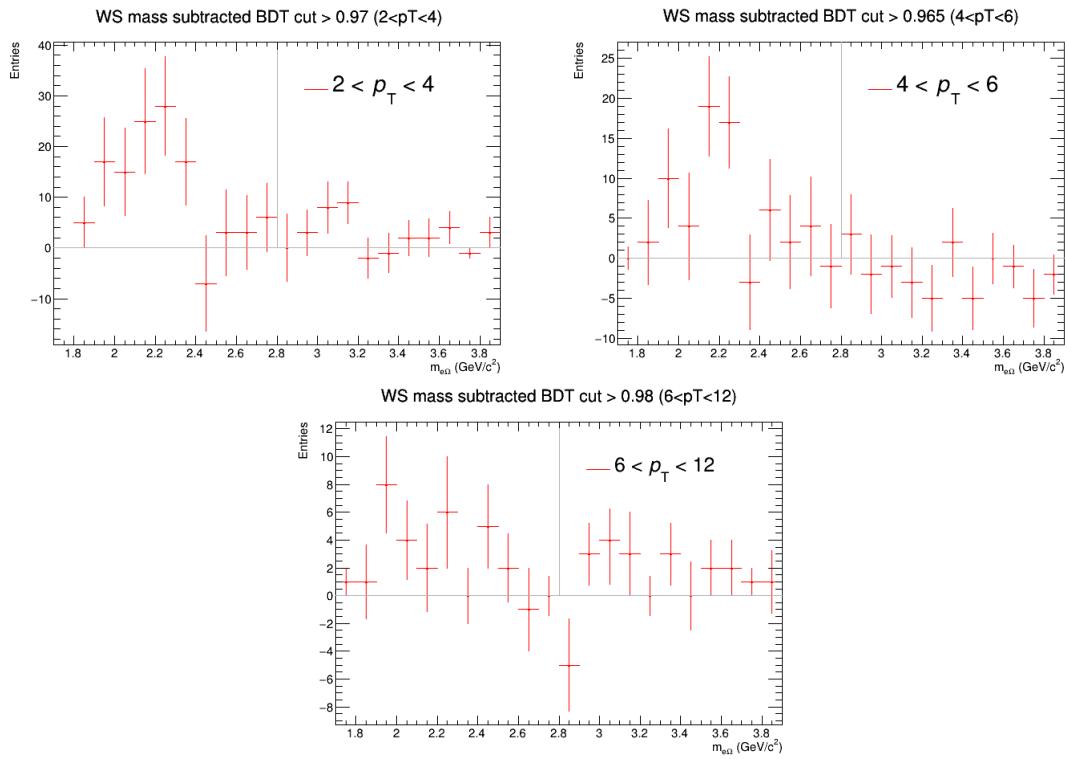


Figure A.12: The invariant-mass distribution of $e\Omega$ pairs after WS subtracting from RS pairs in each p_T interval.

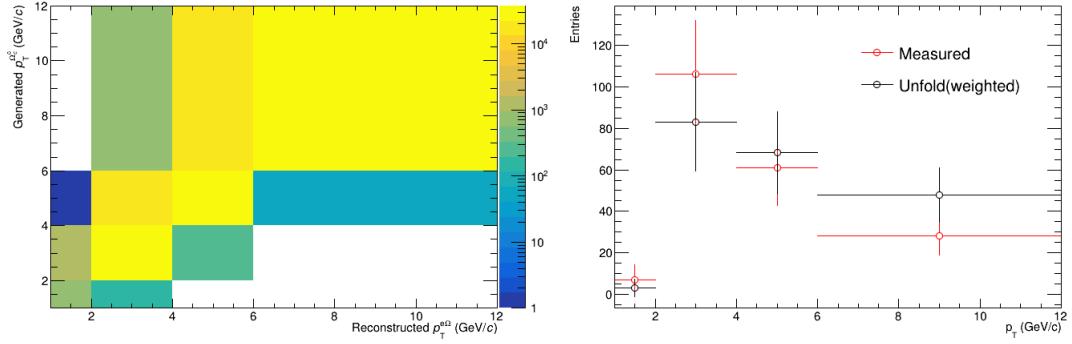


Figure A.13: Left panel: The response matrix between the generated Ω_c^0 -baryon p_T and the reconstructed $e\Omega$ pair p_T . Right panel: The raw yield comparison.

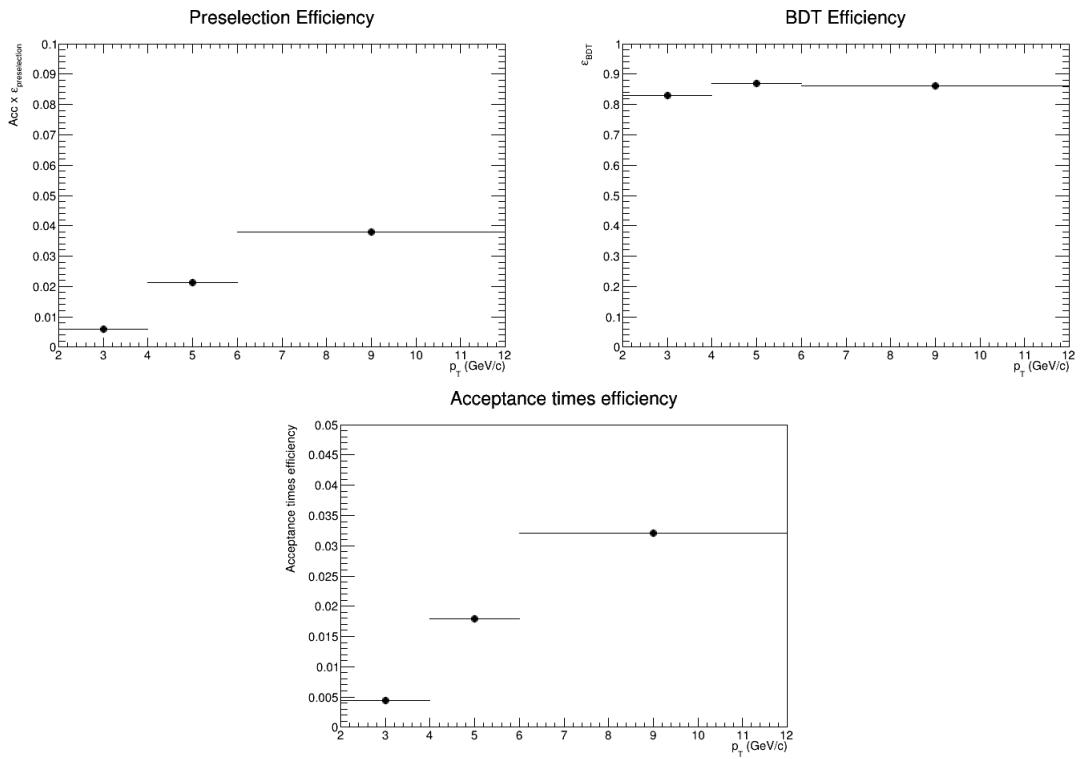


Figure A.14: Top left: Preselection efficiency. Top right: BDT efficiency. Bottom: total acceptance times efficiency for the inclusive Ω_c^0 baryons.

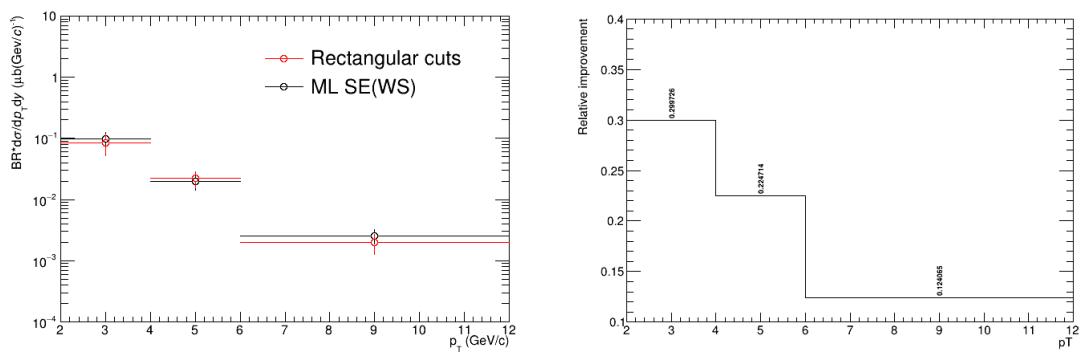


Figure A.15: Left panel: The comparison of p_T -differential production cross section of inclusive Ω_c^0 baryons multiplied by the branching ratio into $\Omega^- e^+ \nu_e$, between the traditional rectangular cut method (in red color) and the new ML method (in black color). Right panel: The improvement of relative statistical uncertainties from ML with respect to the traditional rectangular cut method.



A.2.3 Performance with mixed event technique

The ME RS $e\Omega$ pairs are used for the ML training and background subtraction by default. It shows more statistics than the WS pairs in SE, reported in Section 4.1.3.1. In the following, more comparisons from the usage of ME are shown.

A.2.3.1 Comparison of variables

The comparison of variables between MERS (reported in blue) and SEWS (reported in orange) is shown in Fig. A.16, Fig. A.17, and Fig. A.18 for each p_T interval. The resulting distributions are found to be consistent with the two backgrounds.

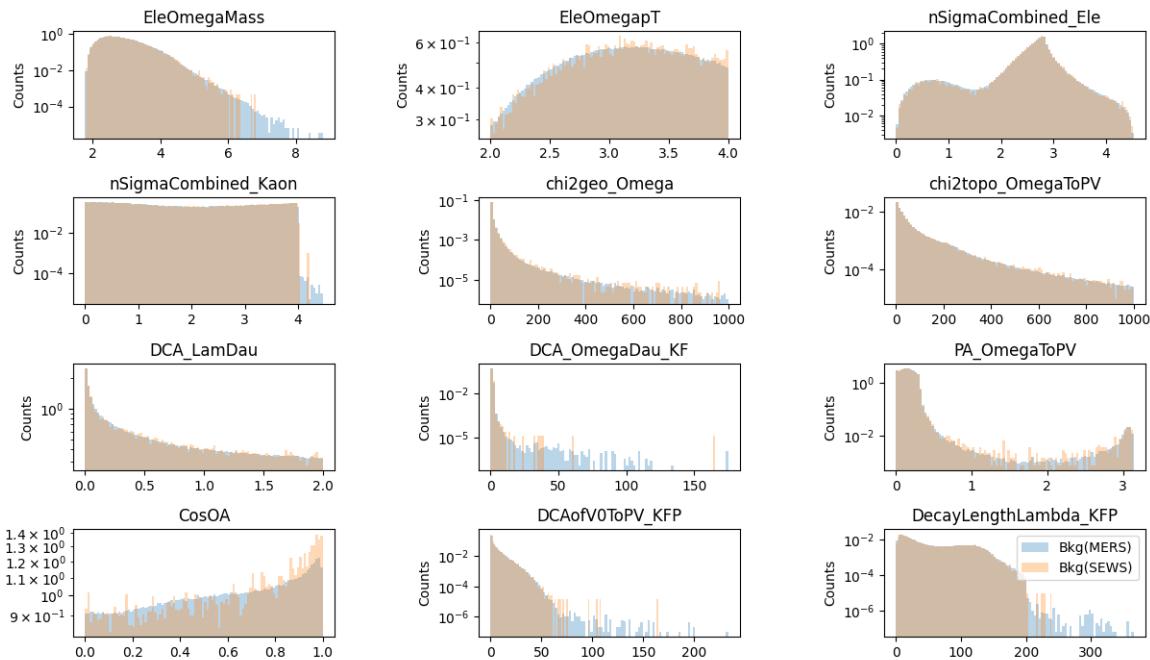


Figure A.16: The comparison of variables between MERS (reported in blue) and SEWS (reported in orange), in $2 < p_T < 4$.

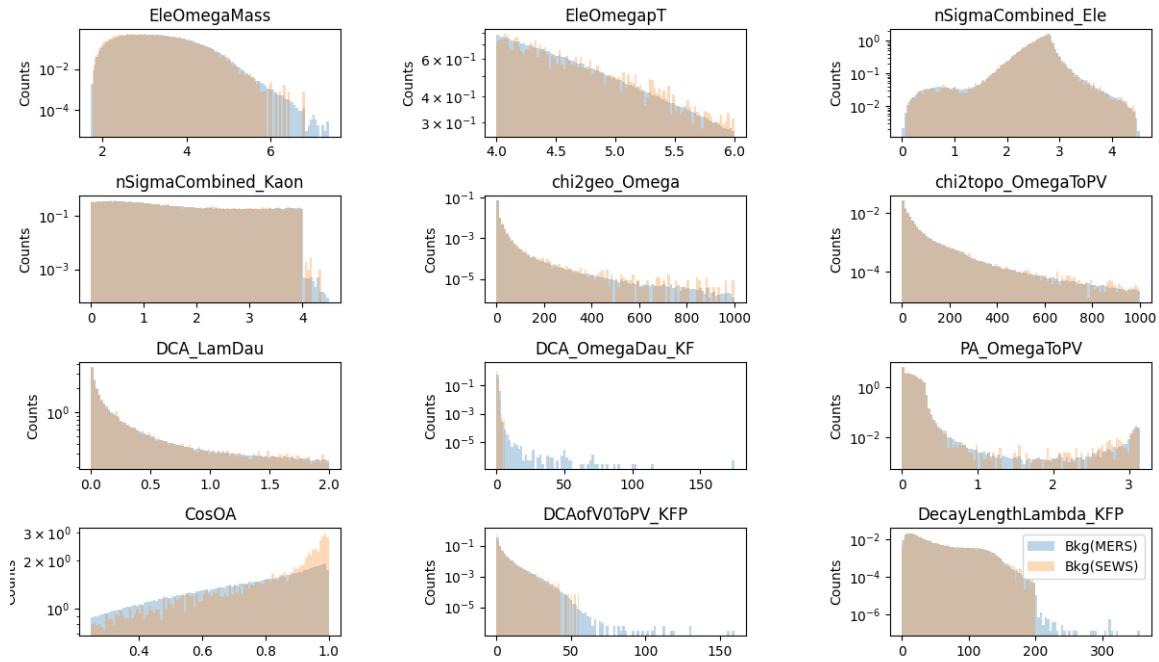


Figure A.17: The comparison of variables between MERS (reported in blue) and SEWS (reported in orange), in $4 < p_T < 6$.

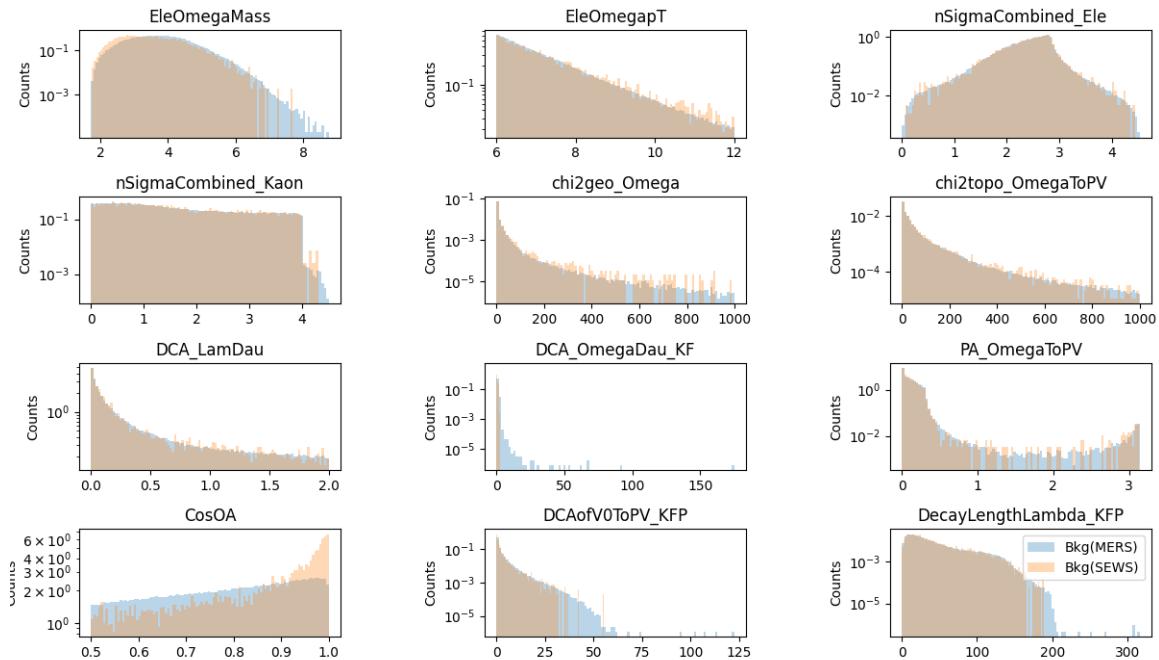


Figure A.18: The comparison of variables between MERS (reported in blue) and SEWS (reported in orange), in $2 < p_T < 4$.



A.2.3.2 Comparison of results between different models

In the following, detailed results obtained from the two different ML models are presented.

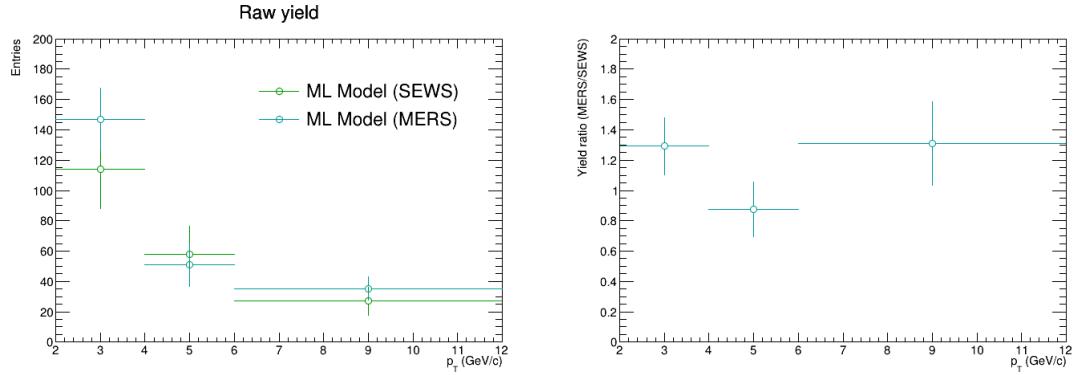


Figure A.19: Left panel: The raw yield comparison from two different ML models: MERS and SEWS. Right panel: The raw yield ratio.

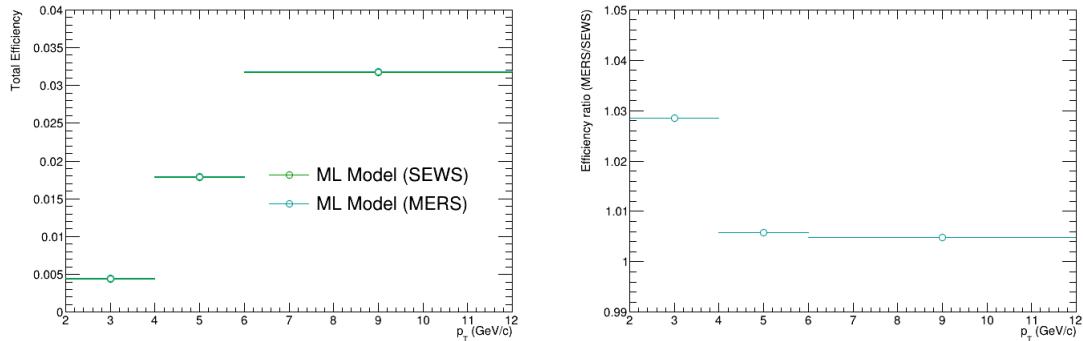


Figure A.20: Left panel: The total reconstruction efficiency comparison from two different ML models: MERS and SEWS. Right panel: The total reconstruction efficiency ratio.

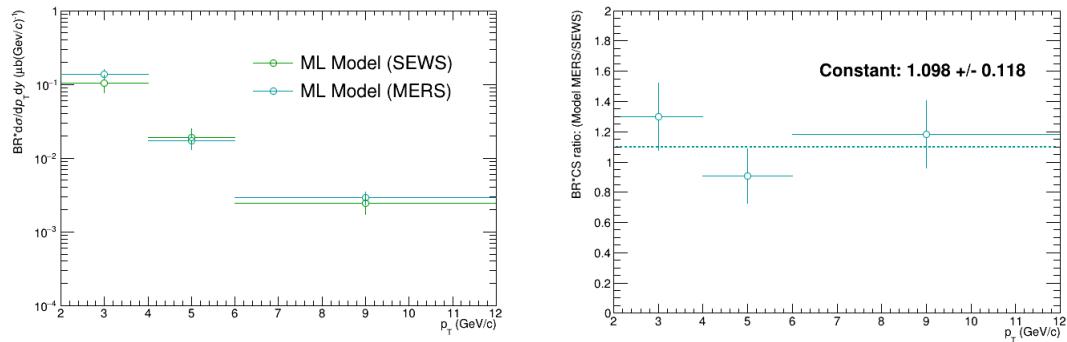


Figure A.21: Left panel: The comparison of p_T -differential production cross section of inclusive Ω_c^0 baryons multiplied by the branching ratio into $\Omega^- e^+ \nu_e$ from two different ML models: MERS and SEWS. Right panel: The p_T -differential production cross section ratio.



Table A.7: The relative statistics error of the p_T -differential production cross section of inclusive Ω_c^0 baryons multiplied by the branching ratio into $\Omega^- e^+ \nu_e$ from two different ML models shown in Fig. A.21, for the p_T intervals $2 < p_T < 4 \text{ GeV}/c$, $4 < p_T < 6 \text{ GeV}/c$, and $6 < p_T < 12 \text{ GeV}/c$.

Relative statistics error ($p_T \text{ GeV}/c$)	2–4	4–6	6–12
ML Model (SEWS)	0.267	0.297	0.293
ML Model (MERS)	0.162	0.244	0.193

Fig. A.19 shows the raw yield comparison obtained from two different ML models of MERS and SEWS being background correspondingly. The MERS gives more yield than SEWS case.

Fig. A.19 shows the reconstruction efficiency comparison obtained from two different ML models of MERS and SEWS being background correspondingly. The ratio is close to unity.

Fig. A.21, the left panel reports the comparison of p_T -differential production cross section of inclusive Ω_c^0 baryons multiplied by the branching ratio into $\Omega^- e^+ \nu_e$ from two different ML models, and their relative statistics errors are shown in Tab. A.7, respectively. As expected, the implementation of ME technique improves slightly the statistical error on the measurement, and the improvement of relative statistics error with respect to the result with SE is: **21%, 25%, 22%** for the p_T intervals $2 < p_T < 4 \text{ GeV}/c$, $4 < p_T < 6 \text{ GeV}/c$, and $6 < p_T < 12 \text{ GeV}/c$. The right panel is the ratio between the ME and SE, fitted by pol0 function, about 10%. It is important to note that the 10% is coming from the yield extraction, since the reconstruction efficiency ratio is close to unity, indicating the advantage of the usage of the ME technique.



A.2.4 Variable distributions and correlation matrix

As mentioned above in Section 4.3.1, the signal and background distribution for $1 < p_T < 2$, $4 < p_T < 6$ and $6 < p_T < 12$ is shown in the following Fig. A.22, Fig. A.23 and Fig. A.24. It's worth noting that the p_T intervals used for the final result are $[1, 2, 4, 6, 12]$. The reason of starting from $[1, 2]$ is to avoid the possible edge effect introduced by the unfolding procedure. The same strategy is adopted in Refs [21, 33].

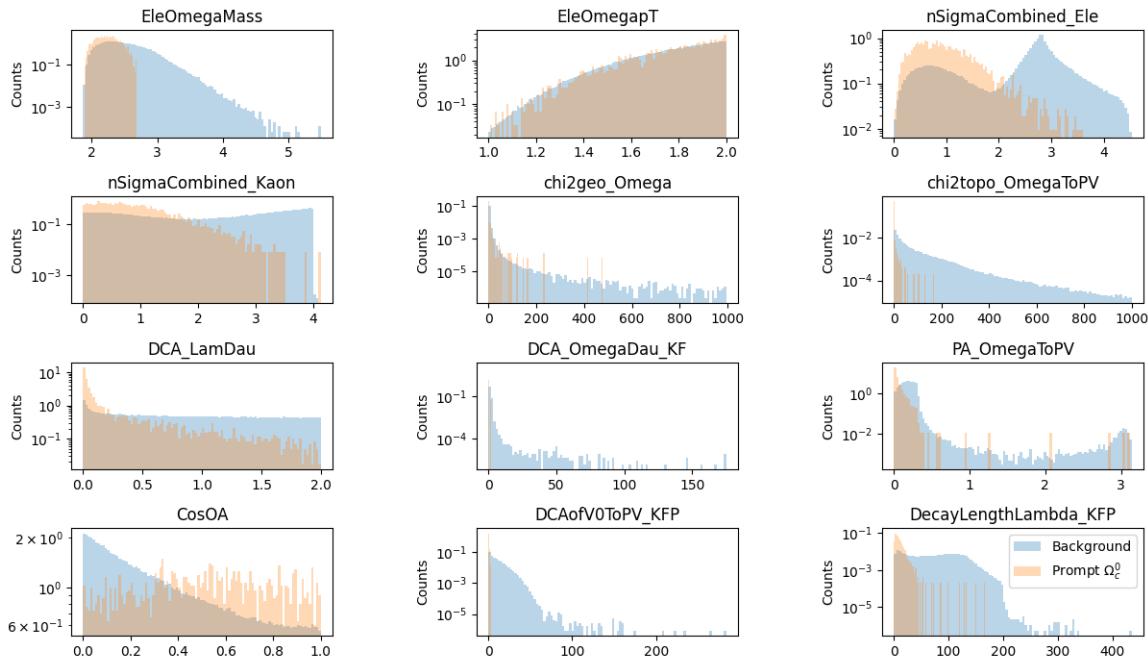


Figure A.22: Training variables (expect EleOmegaMass, EleOmegapT, CosOA) distribution in $1 < p_T < 2$ for signal (reported in orange) and background (reported in blue) candidates.

Fig. A.25, A.26, and A.27 depict the correlation matrices for all selected classification criteria in the signal and the background samples for p_T intervals: $1 < p_T < 2$, $4 < p_T < 6$ and $6 < p_T < 12$.

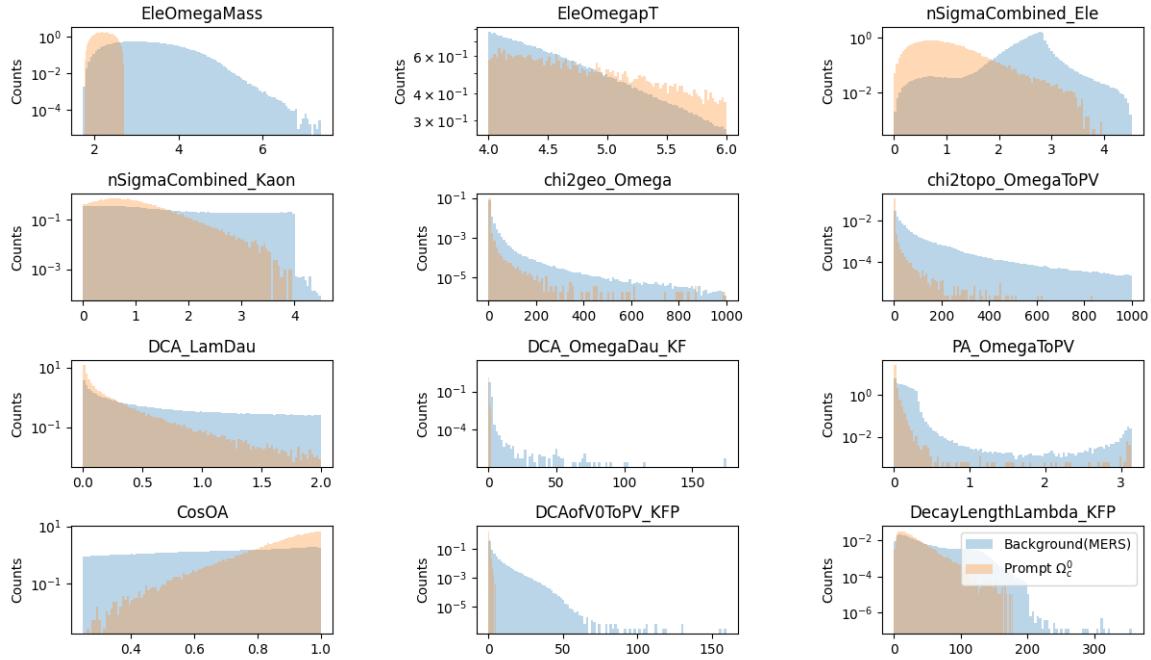


Figure A.23: Training variables (expect EleOmegaMass, EleOmegapT, CosOA) distribution in $4 < p_T < 6$ for signal (reported in orange) and background (reported in blue) candidates.

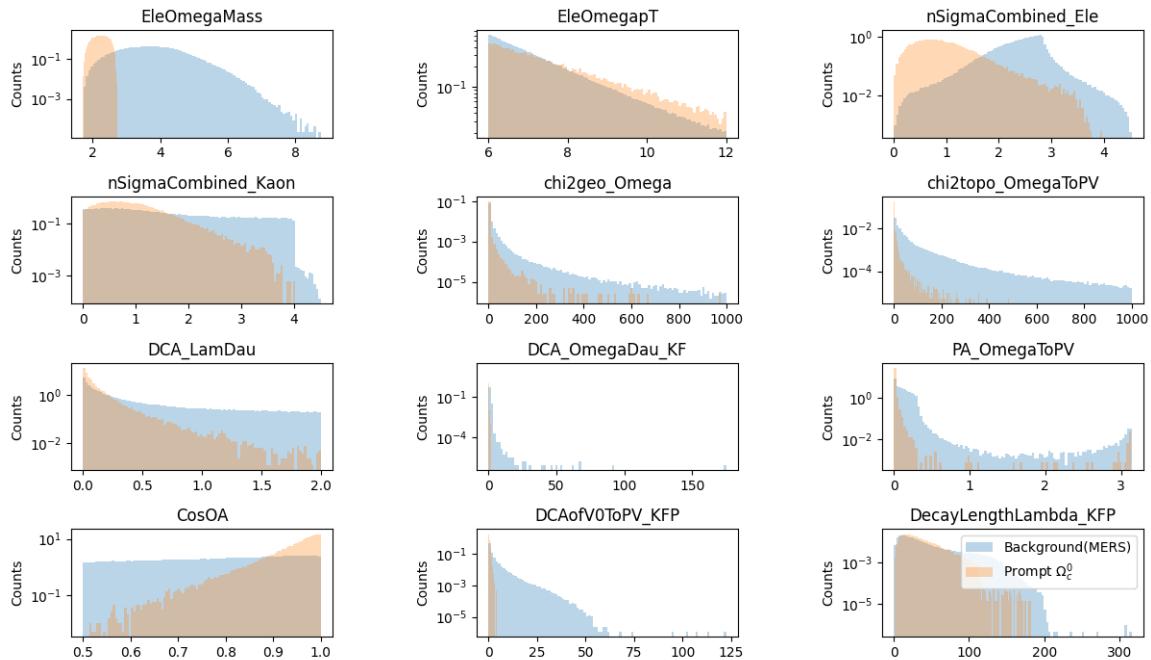


Figure A.24: Training variables (expect EleOmegaMass, EleOmegapT, CosOA) distribution in $6 < p_T < 12$ for signal (reported in orange) and background (reported in blue) candidates.

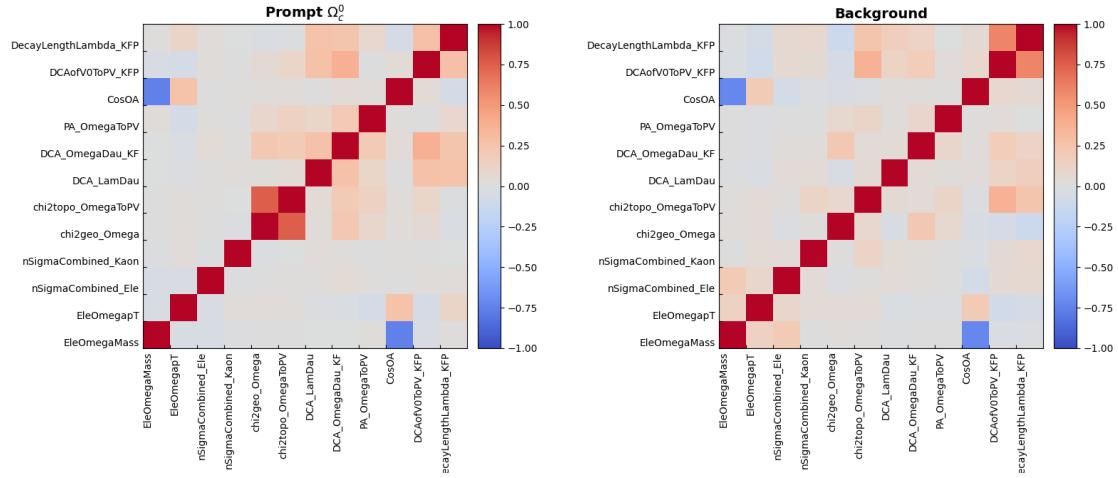


Figure A.25: Correlation matrix in the transverse momentum range $1 < p_T < 2$.

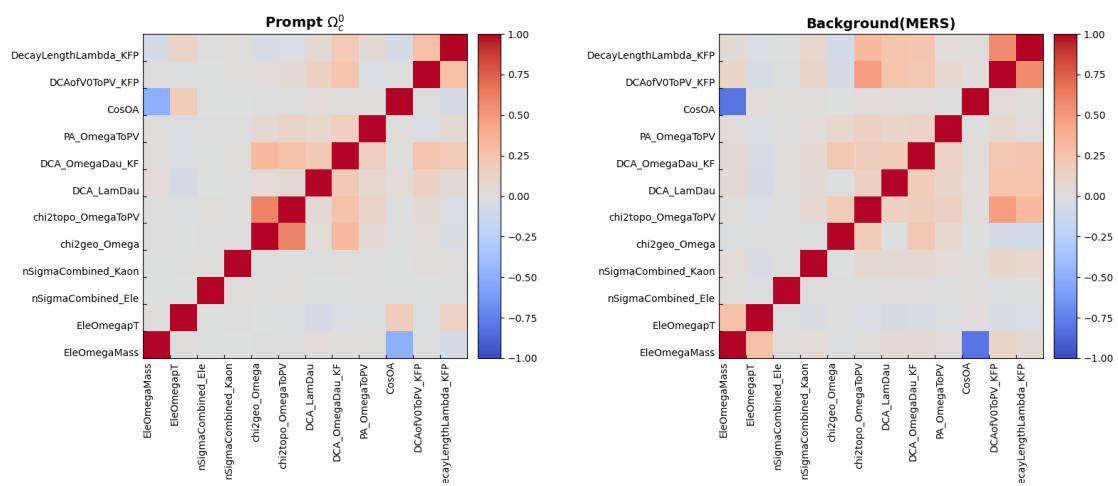


Figure A.26: Correlation matrix in the transverse momentum range $4 < p_T < 6$.

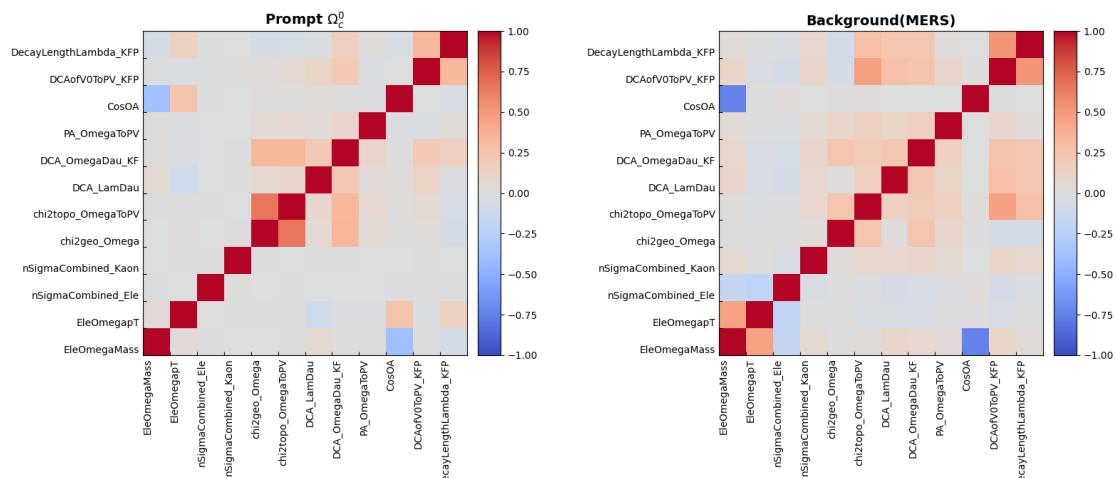


Figure A.27: Correlation matrix in the transverse momentum range $6 < p_T < 12$.



A.2.5 ML output

As mentioned in Section 4.3.3, here Fig. A.28, A.29, A.30 and A.31 are the rest corresponding comparisons.

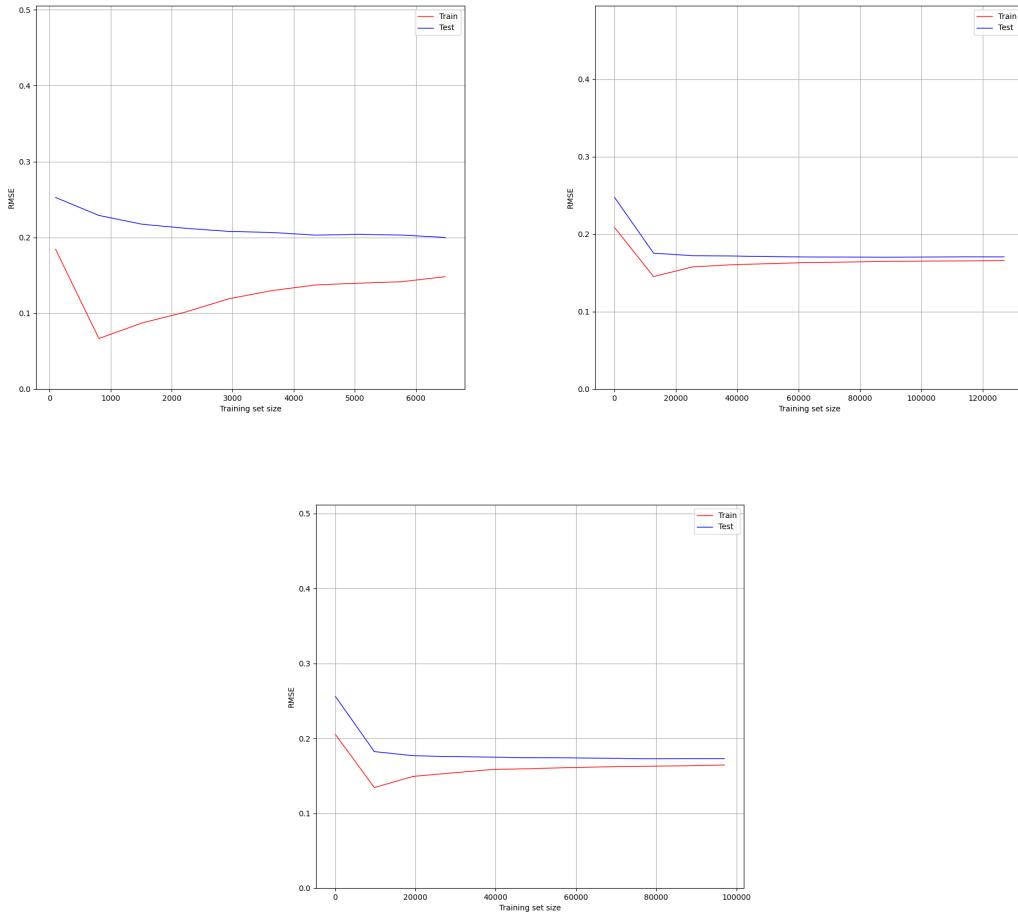


Figure A.28: The learning curves of the training set and test set in each p_T interval in this analysis. Top left panel: $1 < p_T < 2 \text{ GeV}/c$. Top right panel: $4 < p_T < 6 \text{ GeV}/c$. Bottom panel: $6 < p_T < 12 \text{ GeV}/c$.

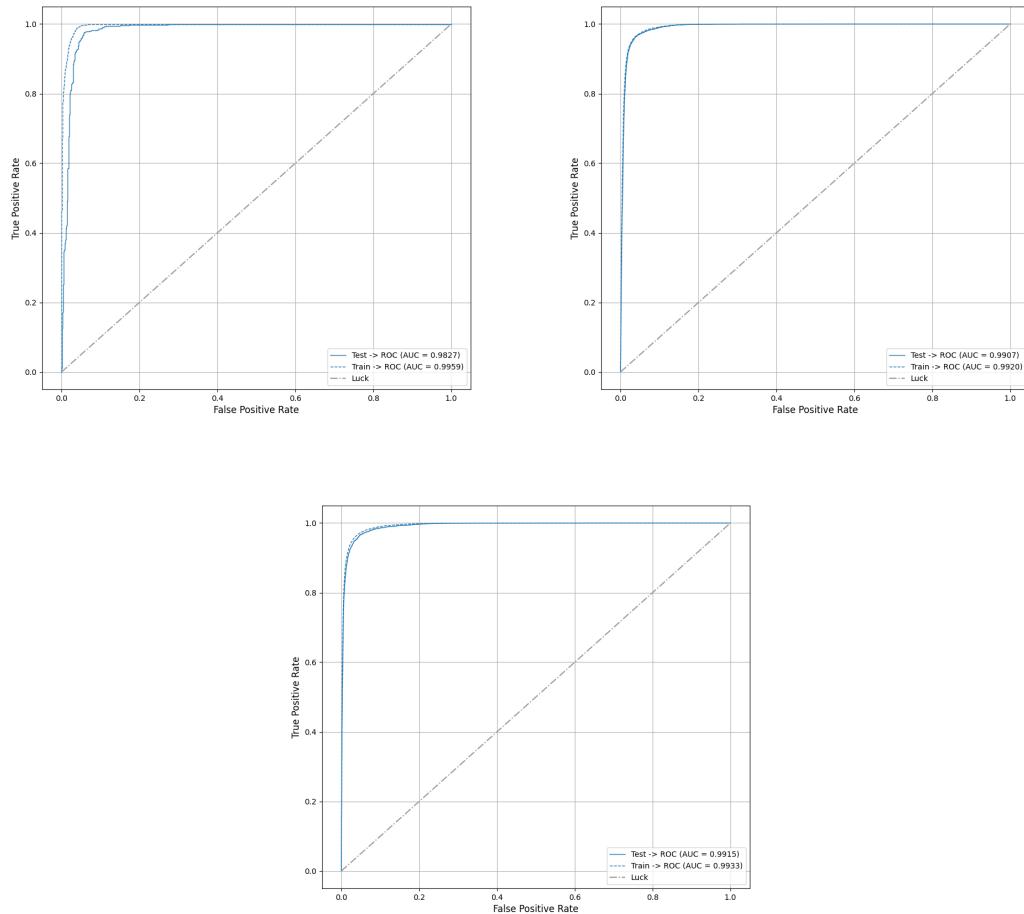


Figure A.29: The ROC-AUC curves for the training set and test set in each p_T interval in this analysis. Top left panel: $1 < p_T < 2 \text{ GeV}/c$. Top right panel: $4 < p_T < 6 \text{ GeV}/c$. Bottom panel: $6 < p_T < 12 \text{ GeV}/c$.

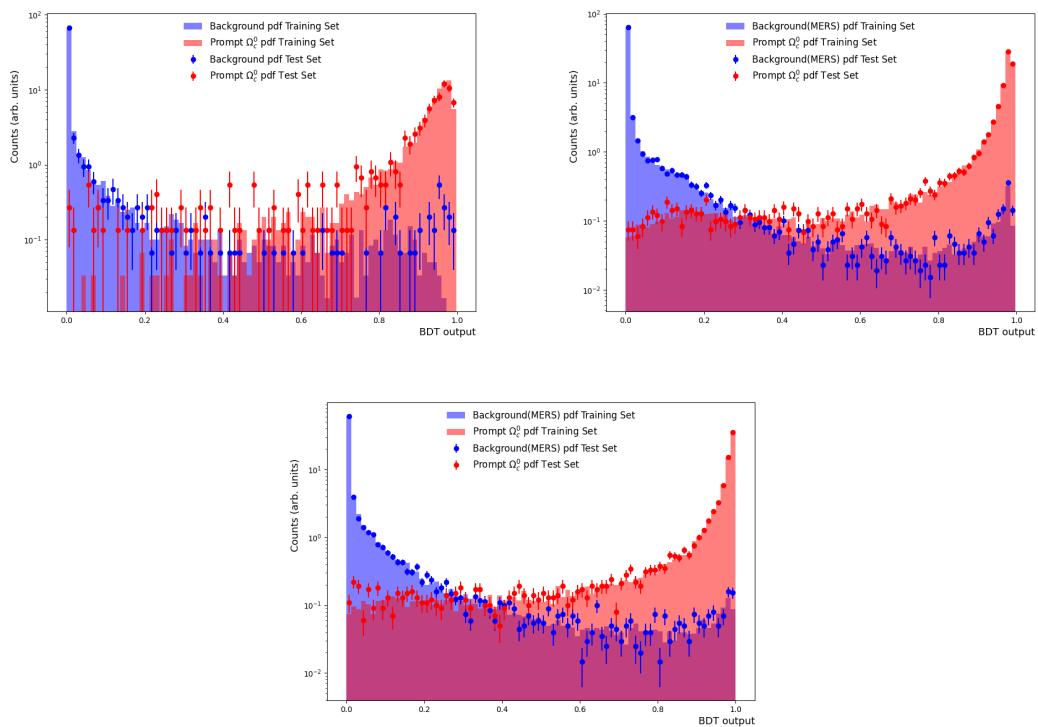


Figure A.30: Machine learning model in each p_T interval in this analysis. Top left panel: $1 < p_T < 2 \text{ GeV}/c$. Top right panel: $4 < p_T < 6 \text{ GeV}/c$. Bottom panel: $6 < p_T < 12 \text{ GeV}/c$.

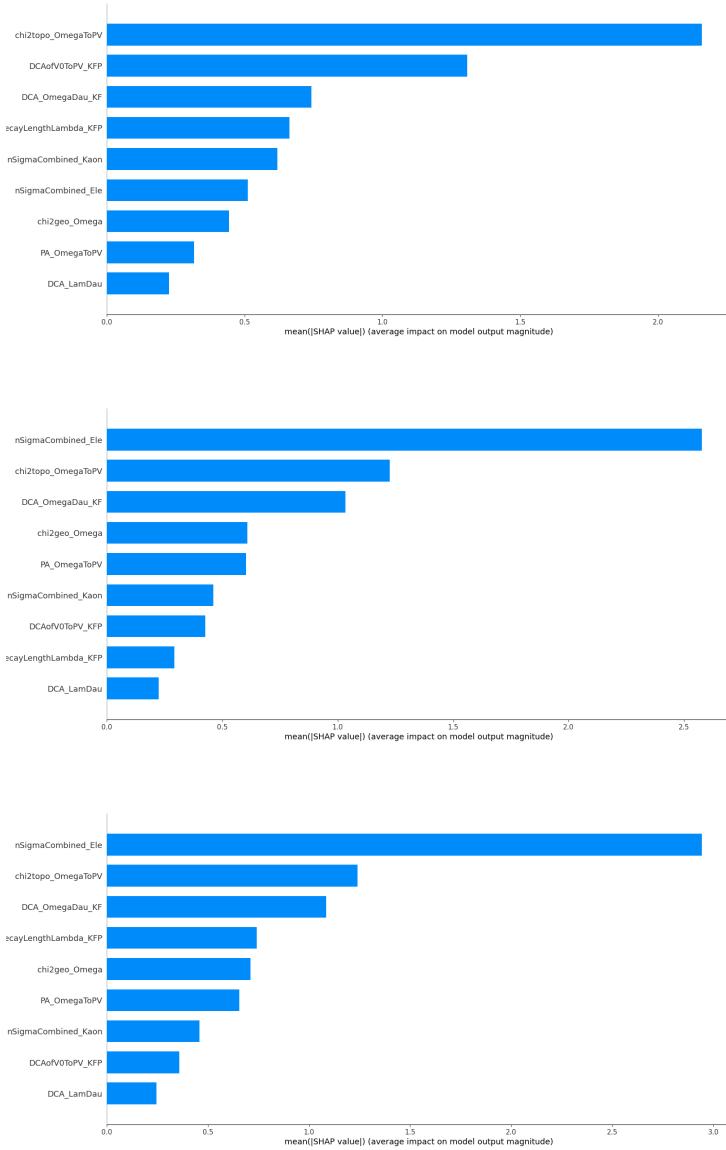


Figure A.31: Feature importance of the classification criteria in each p_T interval in this analysis. From top to bottom: $1 < p_T < 2 \text{ GeV}/c$, $4 < p_T < 6 \text{ GeV}/c$, and $6 < p_T < 12 \text{ GeV}/c$.



A.2.6 Working Point

Following Section 4.3.4, Figs A.32, A.33, A.34 and A.35, report the rest results.

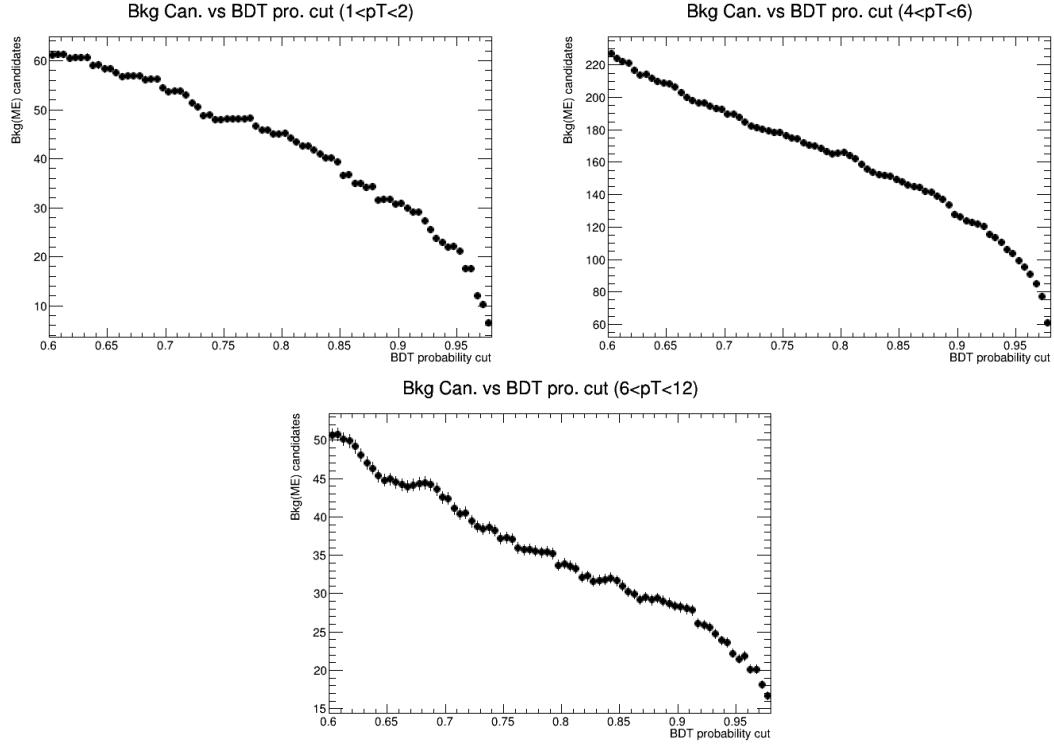


Figure A.32: The number of background candidates (MERS scaled by SEWS) as a function of BDT cut for each p_T interval in this analysis. Top left panel: $1 < p_T < 2 \text{ GeV}/c$. Top right panel: $4 < p_T < 6 \text{ GeV}/c$. Bottom panel: $6 < p_T < 12 \text{ GeV}/c$.

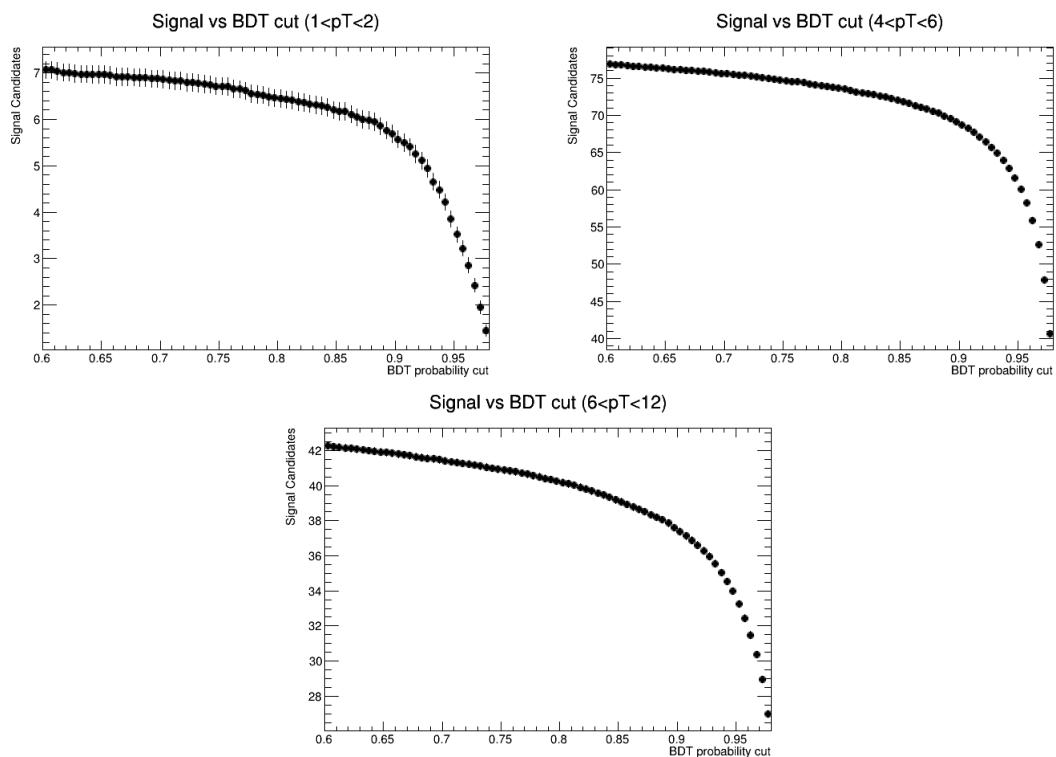


Figure A.33: The number of estimated expected signal candidates as a function of BDT cut for each p_T interval. Top left panel: $1 < p_T < 2$ GeV/c. Top right panel: $4 < p_T < 6$ GeV/c. Bottom panel: $6 < p_T < 12$ GeV/c.

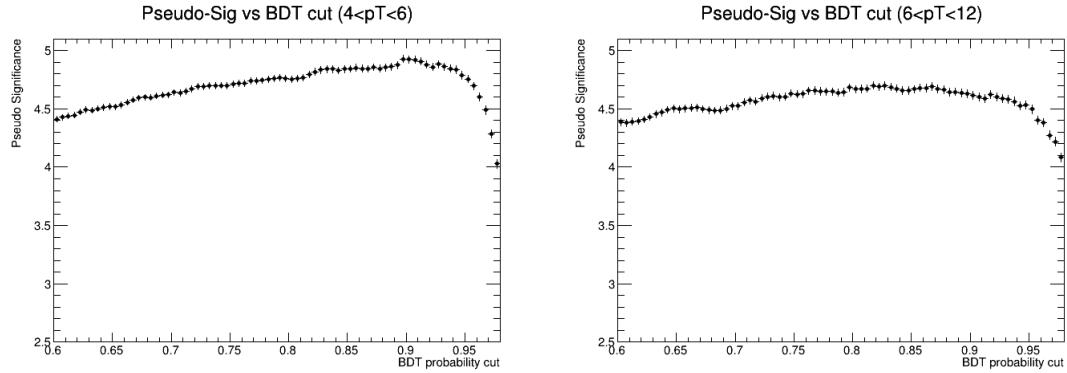


Figure A.34: The pseudo-significance as a function of BDT cut for each p_T interval. Top left panel: $2 < p_T < 4 \text{ GeV}/c$. Top right panel: $4 < p_T < 6 \text{ GeV}/c$. Bottom panel: $6 < p_T < 12 \text{ GeV}/c$.

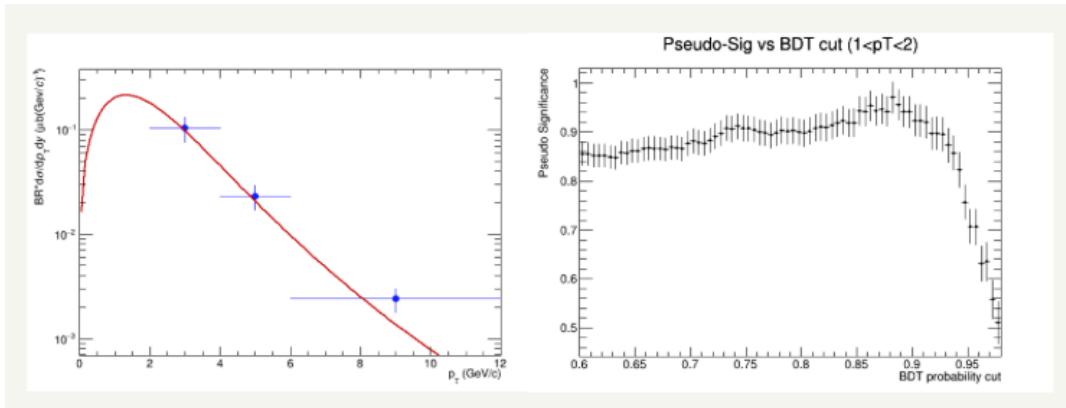


Figure A.35: $1 < p_T < 2$: The Tsallis function is used to fit the result of hadronic Ω_c^0 . The pseudo-significance as a function of BDT cut ($\text{WP}_{\text{BDT}} = 0.95$) to reject backgrounds from other hadrons.



A.2.7 Acceptance efficiency

The $(A \times \epsilon)$ correction factors of prompt, beauty feed-down (non-prompt), and inclusive Ω_c^0 as a function of p_T , shown in Fig. A.36, are observed to be consistent with each other within uncertainties because the selection variables used are not sensitive to the displacement by a few hundred micrometers of the prompt and beauty feed-down Ω_c^0 decay vertices from the collision point.

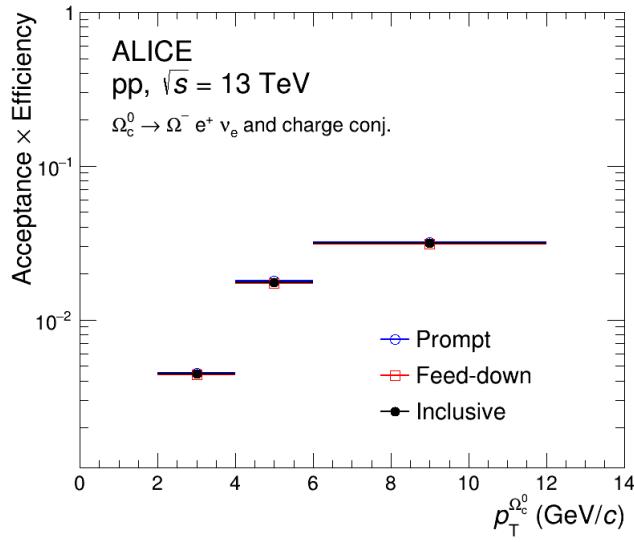


Figure A.36: Product of acceptance and efficiency for prompt, feed-down, and inclusive Ω_c^0 baryons in this analysis.

A.2.8 Systematic: ITS–TPC matching

Same procedure as performed in Appendix A.1.1, two steps are applied to propagate the systematic uncertainty on electron tracks to Ω_c^0 .

Firstly, the systematic uncertainties of ITS–TPC matching efficiency of electrons at the track level are given by DPG group period by period [194]. The average values used from DPG are shown in Tab. A.8.

Table A.8: The systematic uncertainties of ITS–TPC matching efficiency for electron taken from DPG in each p_T bin at $\sqrt{s} = 13$ TeV.

$p_T^{e \rightarrow \Omega_c^0}$	(1, 2)	(2, 3)	(3, 4)	(4, 5)	(5, 6)	(6, 7)	(7, 8)	(8, 9)
ITS–TPC matching	1.8%	2.7%	2.7%	2.3%	2.3%	2.4%	2.4%	3%

$$\text{Syst}_{\text{ITS--TPC matching}}^{e\Omega} = \frac{\sum_{i=1}^N (\text{Syst}_i^{e \leftarrow \Omega_c^0} * \text{BC}_i)}{\sum_{i=1}^N \text{BC}_i}, \quad (\text{A.5})$$

where



- $\text{Syst}_i^{e\Omega}$ is the systematic of ITS–TPC matching in i^{th} $p_T^{e \rightarrow \Omega_c^0}$ bin as shown in Tab. A.8.
- BC_i is the i^{th} bin content in $p_T^{e \rightarrow \Omega_c^0}$ distribution for each $p_T^{e\Omega}$ bin as shown in Fig. A.37

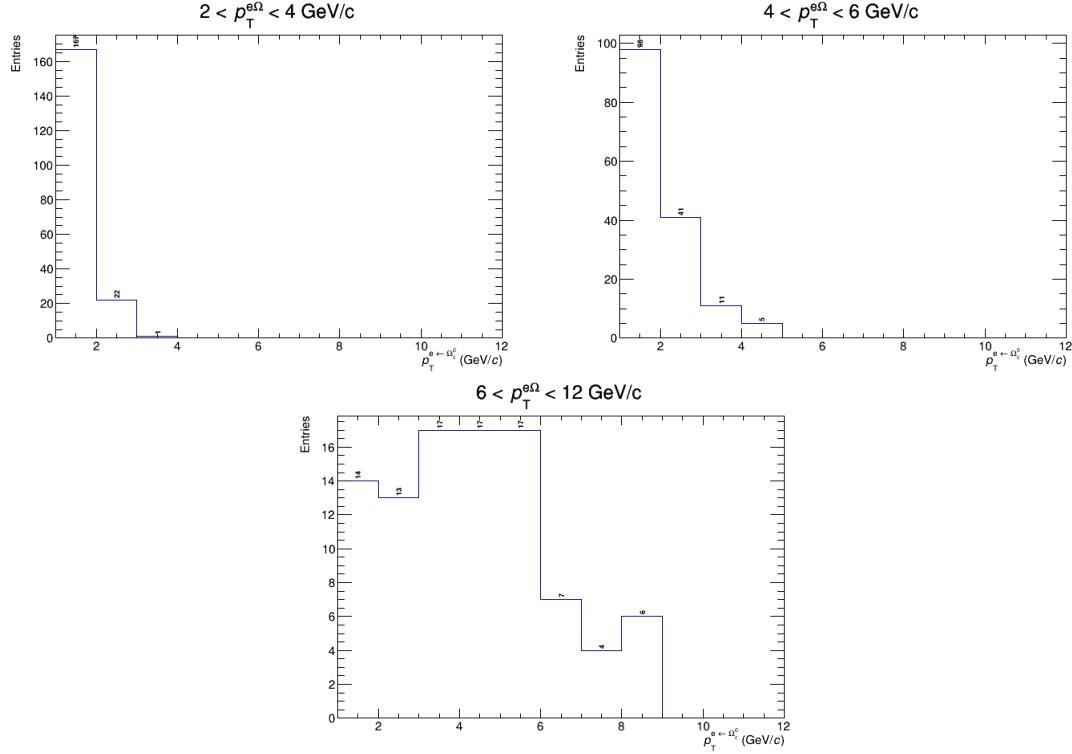


Figure A.37: The distributions of p_T^e in different $p_T^{e\Omega}$ intervals.

$$\text{Syst}_{\text{Bin}[2,4]}^{e\Omega} = \frac{167 * 1.8\% + 22 * 2.7\% + 1 * 2.7\%}{167 + 22 + 1} \approx 1.91\%, \quad (\text{A.6})$$

$$\text{Syst}_{\text{Bin}[4,6]}^{e\Omega} = \frac{98 * 1.8\% + 41 * 2.7\% + 11 * 2.7\% + 5 * 2.3\%}{98 + 41 + 11 + 5} \approx 2.1\%, \quad (\text{A.7})$$

$$\text{Syst}_{\text{Bin}[6,12]}^{e\Omega} = \frac{14 * 1.8\% + 13 * 2.7\% + 17 * 2.7\% + 17 * 2.3\% + 17 * 2.3\% + 7 * 2.4\% + 4 * 2.4\%}{14 + 13 + 17 + 17 + 17 + 7 + 4 + 6} \approx 2\%, \quad (\text{A.8})$$

Table A.9: The systematic uncertainty of $p_T^{e\Omega}$ propagated from p_T^e in this analysis.

$p_T^{e\Omega}$ (GeV/c)	2-4	4-6	6-12
Systematics	1.91%	2.1%	2%

Secondly, propagating the systematic of $p_T^{e\Omega}$ to $p_T^{\Omega_c^0}$, is performed similarly as to the first step. However, as with the missing momentum of neutrino, the response matrix is considered to convert $p_T^{e\Omega}$ to $p_T^{\Omega_c^0}$. To obtain the $p_T^{e\Omega}$ distributions in different $p_T^{\Omega_c^0}$ intervals, the response matrix in the left panel of Fig. 4.20 is projected onto the X-axis. Then the distribution of $p_T^{e\Omega}$ in different $p_T^{\Omega_c^0}$ intervals is shown in Fig. A.38.

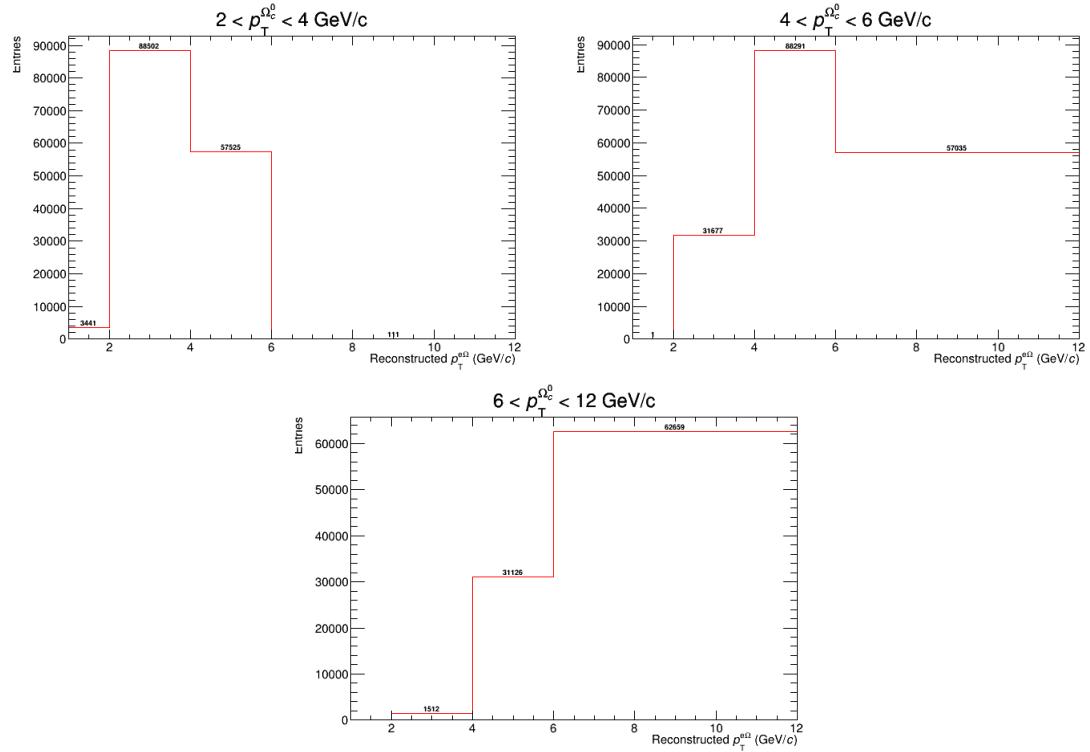


Figure A.38: The distributions of $p_T^{e\Omega}$ in different $p_T^{\Omega_c^0}$ intervals.

The Eq. A.9 is used to propagate the systematic uncertainties of ITS–TPC matching efficiency of $p_T^{e\Omega}$ to $p_T^{\Omega_c^0}$.

$$\text{Syst}_{\text{ITS--TPC matching}}^{p_T^{\Omega_c^0}} = \frac{\sum_{i=1}^N (\text{Syst}_i^{e\Omega \rightarrow \Omega_c^0} * \text{BC}_i)}{\sum_{i=1}^N \text{BC}_i}, \quad (\text{A.9})$$

where

- $\text{Syst}_i^{p_T^{\Omega_c^0}}$ is the systematic of ITS–TPC matching in i^{th} $p_T^{e\rightarrow\Omega_c^0}$ bin as shown in Tab. A.9.
- BC_i is the i^{th} bin content in $p_T^{e\rightarrow\Omega_c^0}$ distribution for each $p_T^{e\Omega}$ bin as shown in Fig. A.38.

$$\text{Syst}_{\text{Bin}[2,4]}^{\Omega_c^0} = \frac{88502 * 1.91\% + 57525 * 2.1\%}{88502 + 57525} \approx 2\%, \quad (\text{A.10})$$

$$\text{Syst}_{\text{Bin}[4,6]}^{\Omega_c^0} = \frac{31677 * 1.91\% + 88291 * 2.1\% + 57035 * 2\%}{31677 + 88291 + 57035} \approx 2\%, \quad (\text{A.11})$$

$$\text{Syst}_{\text{Bin}[6,12]}^{\Omega_c^0} = \frac{1512 * 1.91\% + 31126 * 2.1\% + 62659 * 2\%}{1512 + 31126 + 62659} \approx 2\%, \quad (\text{A.12})$$



博士学位论文
DOCTORAL DISSERTATION



Table A.10: The systematic uncertainty of $p_T^{\Omega_c^0}$ propagated from $p_T^{e\Omega}$.

$p_T^{\Omega_c^0}$ (GeV/c)	2-4	4-6	6-12
Systematic	2%	2%	2%

In the end, 2% is assigned for the final systematic uncertainty of the ITS–TPC matching in this analysis, seen Table. A.10.



A.2.9 Systematic: BDT selection

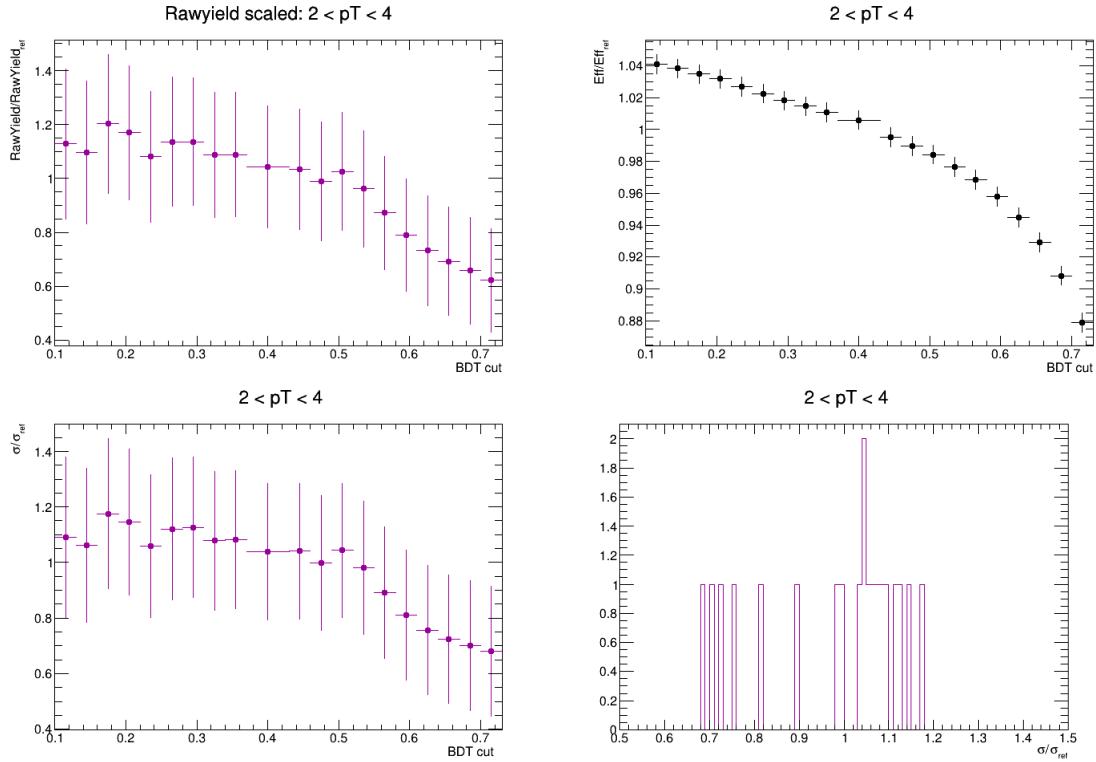


Figure A.39: Top left panel: the ratios of the raw yield as a function of the BDT cuts to the default one. Top right panel: the ratios of the reconstruction efficiency as a function of the BDT cuts to the default one. Bottom left panel: the ratios of the corrected yield as a function of the BDT cuts to the default one. Bottom right panel: the systematic uncertainty, $\sqrt{(\text{RMS})^2 + (\text{shift})^2}$, as a function of BDT in $2 < p_T < 4$.

Table A.11: Electron identification used to study the uncertainties in eID.

Cuts variables	Lower limit of $n\sigma_{\text{TPC}}^e$
VeryLoose	$-4.3 + 1.17 * p_T - 0.094 * p_T^2$
Loose	$-4.1 + 1.17 * p_T - 0.094 * p_T^2$
Standard	$-3.9 + 1.17 * p_T - 0.094 * p_T^2$
Tight	$-3.7 + 1.17 * p_T - 0.094 * p_T^2$
VeryTight	$-3.5 + 1.15 * p_T - 0.09 * p_T^2$

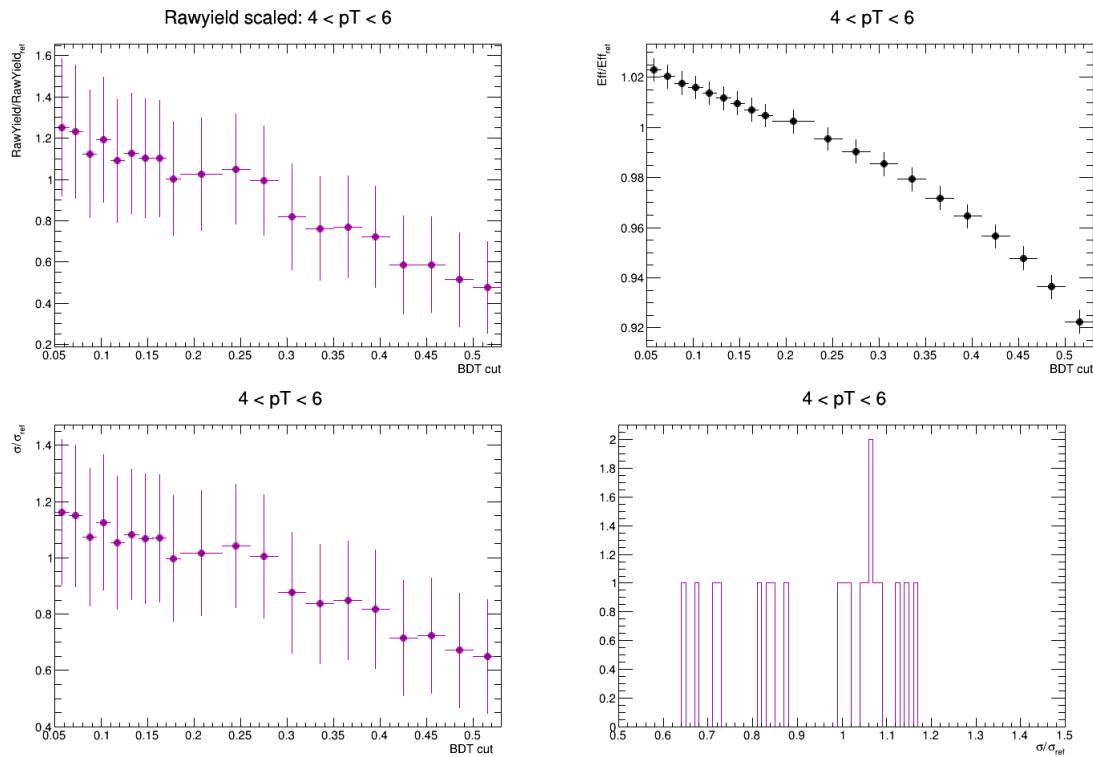


Figure A.40: Top left panel: the ratios of the raw yield as a function of the BDT cuts to the default one. Top right panel: the ratios of the reconstruction efficiency as a function of the BDT cuts to the default one. Bottom left panel: the ratios of the corrected yield as a function of the BDT cuts to the default one. Bottom right panel: the systematic uncertainty, $\sqrt{(\text{RMS})^2 + (\text{shift})^2}$, as a function of BDT in $4 < p_T < 6$.

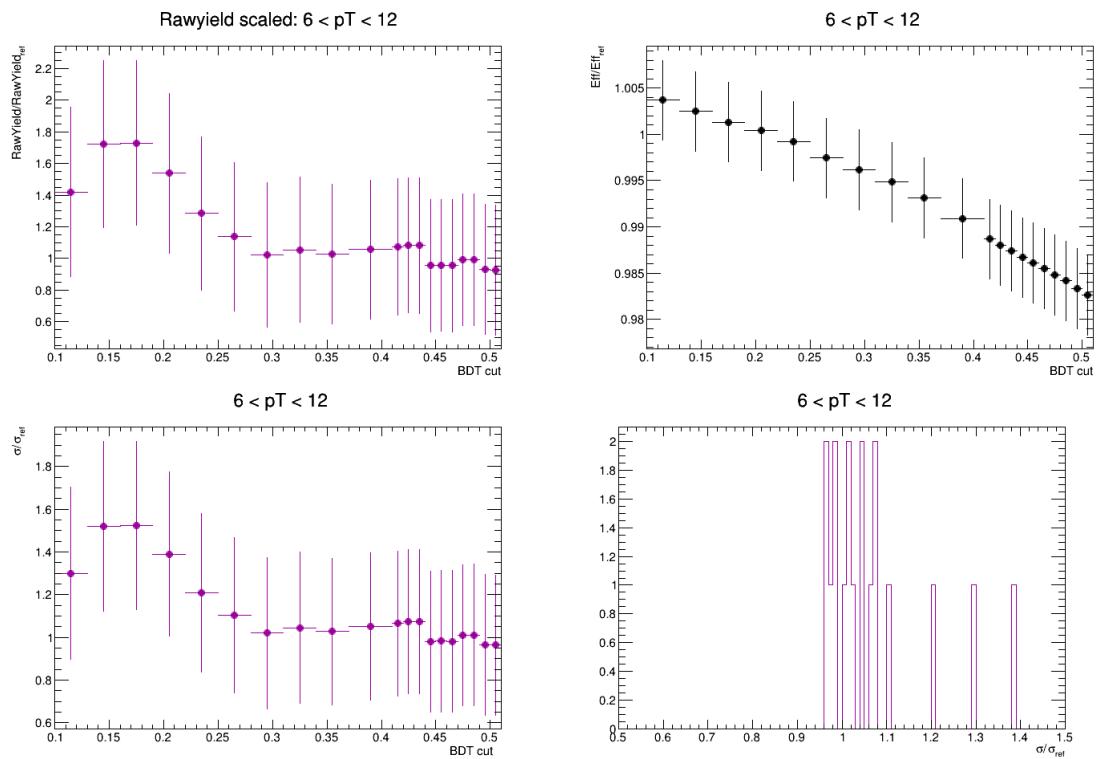


Figure A.41: Top left panel: the ratios of the raw yield as a function of the BDT cuts to the default one. Top right panel: the ratios of the reconstruction efficiency as a function of the BDT cuts to the default one. Bottom left panel: the ratios of the corrected yield as a function of the BDT cuts to the default one. Bottom right panel: the systematic uncertainty, $\sqrt{(\text{RMS})^2 + (\text{shift})^2}$, as a function of BDT in $6 < p_T < 12$.



Appendix B 博士论文中文简介

B.1 论文的研究意义

宇宙大爆炸理论认识在宇宙形成的最初阶段存在一种极其特殊的物质形态—夸克胶子等离子体 (QGP)。为了揭开宇宙的奥妙，在实验室上，大型强子对撞机 (LHC) 中的超相对论重离子碰撞被用来产生 QGP。它是一种解禁闭的状态，即夸克和胶子不被束缚在强子之内。确切地说，LHC 上的大型重离子对撞机实验 (ALICE) 旨在探索 QGP 相变的本质，以及演化过程中 QGP 与硬探针之间的相互作用，进而通过研究这些硬探针来探索 QGP 的相关性质。根据粒子物理标准模型，量子色动力学 (QCD) 是可以描述极端能量密度和温度情况下强相互作用的规范理论，并且预测到 QGP 的存在。当系统温度低于 150 MeV 时候，核物质会经历一个相变的过程：从解禁闭的状态到夸克胶子束缚在一起的禁闭状态。

重离子的演化过程复杂而且比较短暂，主要可以分为：初始状态，预平衡态，QGP 阶段，以及强子化阶段。强相互作用力可以描述 QGP 和所有核物质中的各向性质，这也是 ALICE 实验的主要研究课题之一。现实中，我们只能通过重建并分析碰撞后末态的强子信息来反推 QGP 的信息。其中，重味，包括粲夸克 (c) 和底夸克 (b)，是探索 QGP 的理想硬探针，因为他们在早期碰撞中的硬散射过程产生，并能够经历整个碰撞演化过程。通常，小系统的实验测量可以用来检验微扰量子色动力学 (pQCD) 的计算，并且可以为核核对撞实验提供参考。重味强子的产额可以通过基于 pQCD 的因子化理论通过卷积得来，通常可以分为三个部分：1) 对撞入射质子的部分子分布函数；2) 硬散射过程中的产生截面；3) 碎裂函数，它表示的是特定夸克通过强子化产生强子的概率，通常认为是普适的，即该过程不依赖于碰撞系统和能量。其中，第一和第三项是非微扰过程，只能通过实验测量来约束，而第二项可以通过 pQCD 来描述。强子化机制可以通过不同种类强子产额的比值来学习，因为部分子分布函数和部分子散射截面是普适的，从而在计算比值时候可以相抵消。

长期以来，基于正负电子 (e^+e^-) 和电子-质子 (ep) 对撞实验的测量数据，夸克碎裂函数一直被认为是普适得。近几年来，ALICE 实验组发表了很多粲介子和重子的产额结果，以及不同的重子-介子产额比值。那些模型基于 pQCD 计算通过正负电子 (e^+e^-) 和电子-质子 (ep) 提供的参数，可以成功地描述 ALICE 的粲介子数据。但是他们却都低估了近期测量的粲重子-介子产额的比值，暗示着或许我们应该重新审视一直所认为的碎裂化函数的普适性。从实验数据上来说，首先有 Λ_c^+/D^0 ，不仅在质子-质子碰撞中观察到在低横动量区间显著高于基于 e^+e^- 碎裂函数参数化的事件产生器 PYTHIA8 的模拟结果，而且，在 ALICE 的中心快度区以及 LHCb 的向前区的质子-铅核 (p-Pb) 对撞中也有类似的现象。这些明显的差别引起了理论家们的注意力，因而一



些改善的模型应运而出，其中包括：1) 考虑那些在夸克胶子等离子体系统内可能包含的动力过程，即 colour reconnection beyond leading colour, quark recombination, Catania; 2) 将强子化过程考虑为统计过程，并将那些并未观测到的高质量的粲强子态也计算进入，即 statistic hadronisation model included relativistic quark model (SHM+RQM)。这些模型成功地描述 Λ_c^+/D^0 数据。另外一个就是 ALICE 的 $\Sigma_c^{++,0}/D^0$ 结果，同样相比 e^+e^- 的数据，其数值明显偏大，另一方面也可以部分解释 Λ_c^+/D^0 的数值偏高是来源于 Σ_c 的非瞬时衰变。这几个新的模型也可以描述该实验结果。再者 ALICE 中还有 $\Xi_c^{+,0}/D^0$ 的测量结果，值得注意的是，这些新模型没有一个能够描述此结果，理论数值都低于实验数据。相比较下，只有 Catania 模型能够预测出数据分布特性。这个新的现象不得不引发我们的思考，考虑这些强子的夸克组成部分， Λ_c^+ (uds), $\Sigma_c^{++,0}$ (uuc, ddc), $\Xi_c^{+,0}$ (usc, dsc)。当强子同时含有粲、奇异夸克时候，模型预测的数值低于重子-介子比数值，相比 Λ_c^+ 和 $\Sigma_c^{++,0}$, $\Xi_c^{+,0}$ 对于理论模型具有更多的束缚力。当考虑更重的粲奇异强子， Ω_c^0 ，这是 ALICE 最新的实验数据，但是由于目前还没有 Ω_c^0 在其强子衰变道 ($\Omega^- \pi^+$) 绝对分支比的数值，因此最终结果以散射截面乘以其分支比来呈现 $(BR \times \Omega_c^0)/D^0$ 。多个理论数据用来计算出该分支比的理论中心值，以及系统误差数值。上面提到的几个模型结果分别乘以该理论分支比数值，但是由于比较大的系统误差，因为我们目前还无法得出很明确的结论，但是仍然可以看到，这些模型的预测远远低于实验数据，说明该重子 Ω_c 的产额增强更显著。这些实验观测结果挑战着我们目前对于 QGP 产生的认知，因此需要更多的实验数据来进一步探索潜在的物理机制。

包含奇异夸克的重味强子的产生是重离子碰撞实验中一个引人注目的研究课题。因为重夸克可能与热密物质中的其他介质成分组合为强子，与小系统相比，热密介质中奇异夸克对会通过热产生而增多，夸克重新组合机制可能会导致含有奇异夸克的重味强子的产额比那些不含奇异夸克的重味强子要高。该论文旨在学习粲奇异强子在其半轻子衰变道下的产额研究，具体分别是， $\Xi_c^0 \rightarrow \Xi^- e^+ \nu_e$ 和 $\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e$ ，其中， Ξ_c^0 强子含有一个奇异夸克， Ω_c^0 强子含有两个夸克，并探究粲夸克的强子化机制。

B.2 实验分析方法

由于重味强子的平均衰变长度很短，最大约为几百微米，因此在探测器检测到之前就发生了衰变。在寻找稀有信号时候，尤其是三级重子衰变，在重建粒子时候常常面临的挑战是如何扣除大量的组合背景。好在 ALICE 实验探测器具有很好的粒子鉴别能力，尤其是在很低的横动量区间，这为我们保证了研究对象的可靠性。实验得到的数据需要经过复杂的计算和分析流程，因此需要依赖于先进的计算技术和高效性。实验离线分析的软件是基于 CERN ROOT, 即 AliRoot 和 AliPhysics。除了真实的实验数据之外，蒙塔卡罗 (MC) PYTHIA 模拟产生器用来模拟纸质子-质子碰撞事件。这些产生器产生的粒子会通过 GEANT 模拟的探测器进行运输，最终得到模拟数据，用来物理分析数据的修正。



分析的流程大概可以分为：事件的筛选、径迹的筛选、研究对象的筛选、信号的提取、Unfolding的应用、重建效率的修正、系统误差的学习、最终得到修正后的粒子产额。

B.2.1 Ξ_c^0 分析方法

Ξ_c^0 课题采用的是传统的分析方法，即粒子的筛选和重建根据 Ξ_c^0 的拓扑衰变结构和运动学特征分析，要满足相应的筛选阈值，目的是尽量保证最好的横动量分辨率，高的径迹少选效率，重建真实的径迹，并要避免二级衰变粒子的干扰。对于 electron 对象，确保其纯度 (purity) 是至关重要的一步，可以检验我们的筛选阈值。对于 Ξ ，我们可以得到其在不同横动量区间的不变质量分布图，以此查看阈值的应用。确定这两个子粒子后，接下来就是信号的提取，即配对。值得注意的是，由于实验上测量不到中子的横动量，因此重建配对的时候只是 $e\Xi$ 配对。该分析采用了电荷共轭模式 (charge conjugate mode): 电子和 Ξ 拥有相反的电荷 ($e^\pm \Xi^\mp$)，称之为正确的信号 (Right Sign, RS); 电子和 Ξ 拥有相同的电荷 ($e^\pm \Xi^\pm$)，称之为错误的信号 (Wrong Sign, WS)。RS 对只来源于母粒子 Ξ_c^0 的衰变以及组合背景，而 WS 对只来源于组合背景。这些 $e\Xi$ 对需要满足以下两个条件: $m_{e\Xi} < 2.5 \text{ GeV}/c^2$, e 和 Ξ 径迹之间的余切角要小于 90° 。初步我们可以到的从 Ξ_c^0 衰变出的产额，实际上是 $e\Xi$ 即 RS-WS。之后 Unfolding 技术来修正缺失的中微子，本分析用到的是贝叶斯方法，其默认迭代次数是 3，可以将 $p_T^{e\Xi}$ 转化为 $p_T^{\Xi_c^0}$ 。在该方法中，响应矩阵 (response matrix) 扮演着重要的角色，反映了 MC 产生 (generated, $p_T^{\Xi_c^0}$) 与重建层面 (reconstructed, $p_T^{e\Xi}$) 的横动量分布的二维关系。考虑到探测器的有效性，即无法测量到真实碰撞后产生的 Ξ_c^0 粒子，分析中往往会有个重建效率 ($A \times \varepsilon$, the product of acceptance and efficiency for Ξ_c^0)。最终我们得到了修正后的 Ξ_c^0 粒子的产额。

B.2.2 Ω_c^0 分析方法

和 Ξ_c^0 分析相比， Ω_c^0 课题引入了三个新颖点：1) 卡尔曼滤波法 (Kalman Filter Particle package) 来重建粒子 Ω ; 2) 利用机器学习 (machine learning) 来分离背景和信号; 3) 用混合事件 (mixed-event) 作背景，用于机器学习的训练，和背景扣除。该分析前期做了很多准备：1) 代码的实施，具体是 KF package 以及混合事件的应用；2) quality control (QA) 检查。代码实施完后，需要进行彻底的研究变量对比，主要是研究对象变量分布图的检查 (comparisons of variable distribution)，其中又包含两小部分，首先检查实施 KF package 前后的对比，然后是混合事件和同一事件的变量对比。接下来就是我们的机器学习的应用，采取的是基于 BDT 的梯度决策树算法 XGBoost。决策树 (BDT) 具有简单的机制，并且能够高效地分离样本和背景。

和传统方法不一样的是，起初的阈值筛选是相对比较宽松的，以此初步去掉部分组合背景，为下一步做准备。机器学习训练集是信号 (MC reconstructed) 和背景 (mixed-event right sign, MERS)。首先确定重要性比较大的研究变量，以此用于训练过程，从而得到稳定并高效的模型结果 (BDT)



probability output)。然后将模型分别应用到同一事件的 RS e Ω 对, WS e Ω 对, 混合事件的 RS e Ω 对, 以及 MC e Ω 对。之后通过研究赝显著性 (pseudo-significance) 和 BDT 之间的关系来决定筛选阈值。该筛选阈值的作用和传统方法的筛选阈值 (是一个阈值集) 结果起到相当的效果。由于混合事件的产生是随机且大量的, 因此我们需要将得到的混合事件的 RS e Ω 对归一化到同一事件的 WS e Ω 对, 用来得到信号: SERS - MERS。接下来的步骤和 Ξ_c^0 分析相同, 通过 Unfolding Bayesian 贝叶斯迭代方法来矫正缺失的中微子横动量, 从而得到 $p_T^{\Omega_c^0}$ 。

B.3 实验结果

第三章展示了在 ALICE 质心能量为 5.02 TeV 质子质子碰撞对撞数据得到的的 Ξ_c^0 产额。该结果和已经发表两个结果相比, 相同碰撞体系, 不同的质心能量, 分别是 7 TeV 和 13 TeV, 结果表明粲奇异重子的产生随着质心能量的增大而增加。三种不同质心能量下的重子介子产额的比值 Ξ_c^0/D^0 也进行了比较, 结果显示有一定横动量的依赖性。不同模型和实验 Ξ_c^0 的产额进行对比: 夸克重组模型 QCM, PYTHIA 8 基于碎裂化函数普适的模型 (PYTHIA 8 Monash, 和 PYTHIA 色重组模型 (PYTHIA 8 Mode 0, 2, 3, 进行对比, 这些模型预测数值都低于实验。对于 Ξ_c^0/D^0 比值, 另外有两个新的模型: Catania model 和 SHM+RQM。这些所有模型预测的比值也都低于实验数值。其中只有 Catania 模型能够给出数据分布趋势, 且相对接近实验值。

目前, 理解更重粲奇异强子 Ω_c^0 产生的测量, 最大的一个问题是没有一个精确的绝对分支比数值。所以, 第四章最后的结果是 Ω_c^0 散射截面乘以其半轻子衰变道的分支比。由于引入机器学习方法, 以及混合事件作为组合背景, 背景扣除有明显的提高, 最终结果的统计误差有很大的改善, 并且产额提高 10%。与之前发表的 Ω_c^0 在其强子化的结果, 我们可以得到这两个衰变道的 Ω_c^0 分支比的比值, $BR(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e)/BR(\Omega_c^0 \rightarrow \Omega^- \pi^+) = 1.12 \pm 0.22 \text{ (stat.)} \pm 0.27 \text{ (syst.)}$ 。该结果和三个理论结果以及 CLEO 合作组的数据相吻合, 但是却低于 Belle 合作组的结果, 并相差 2.3 标准差。值得注意的是, 该 Ω_c^0 分支比的比值和之前 ALICE 发表的 $BR(\Xi_c^0 \rightarrow \Xi^- e^+ \nu_e)/BR(\Xi_c^0 \rightarrow \Xi^- \pi^+)$ 数值相兼容。

B.4 总结与展望

该论文旨在通过分析 ALICE Run2 数据来探索粲奇异重子的强子化机制。那些基于碎裂化普适的模型都无法描述 Ξ_c^0 的产额和 Ξ_c^0/D^0 重子介子比值, 再次表明粲碎裂化函数的非普适性, 以及粲重子的增强性。另外, 一些新提出的考虑夸克胶子的动力过程的模型, (例如 CR model, QCM, 或者 SHM+RQM) 也都无法很好的吻合实验数据, 不过 Catania model (假设强子化过程中不仅有碎裂化过程, 也有重组过程) 相对接近实验数据, 这意味着重组机制在强子化过程中起着主要作用。此外, 这些实验数据和理论数值的差异, 有力说明粲异重子对于理论模型有更好的约束力。



目前我们对高能重味夸克的强子化过程还没有透彻的了解。因此，该论文的两个分析对粲奇异重子强子化机制提供了有价值的信息，探索了奇异夸克或奇异双夸克的作用。而且，对针对于粲奇异重子的色重组理论的探索提供更多的参考价值，也帮助探索重组在粲奇异重子强子化的角色。另外，该分析结果可以为理论计算提供进一步的约束，并为在重离子碰撞系统下产生的粲夸克和底夸克与热密介质的相互作用提供了参考。

最后，该分析中在小型碰撞系统下重子产生增强的发现，还有其它小系统系统中的观察到的结果与 Pb-Pb 碰撞中有类似的特征，例如集体模式和奇异性增强，对目前大家只认为 QGP 仅在核-核碰撞形成的观点提出了质疑。更深入地理解粲夸克的强子化机制和强子碰撞过程中的动力学机制需建立在更多的实验数据基础上。基于全新升级内部径迹探测器的 ALICE 第三次运行 Run 3 收集的数据，一些新的实验测量有望实现：

- 多粲强子态的测量，比如 Ξ_{cc}^+ ($cc\bar{d}$), Ξ_{cc}^{++} ($cc\bar{u}$), Ω_{cc}^+ ($cc\bar{s}$), Ω_{ccc}^{++} ($ccc\bar{s}$)
- 学习不同快度区间下的重子介子比
- 底强子的测量，比如 Λ_b , Ξ_b , Ω_b
- 小系统下的集体流效应
- 高质量强子态的研究



攻读学位期间的成果

- **Paper publication: Measurements of the production cross section of prompt Ξ_c^0 baryons at midrapidity in pp collisions at $\sqrt{s} = 5.02$ TeV**
ALICE Collaboration, PC members: T.Cheng (chair), Andrea Dubla, Jianhui Zhu
Journal of High Energy Physics 10 (2021) 159
Paper publication: Measurement of Ω_c^0 baryon production and branching-fraction ratio $\text{BR}(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e)/\text{BR}(\Omega_c^0 \rightarrow \Omega^- \pi^+)$ in pp collisions at $\sqrt{s} = 13$ TeV
ALICE Collaboration, PC members: T.Cheng (chair), Andrea Dubla, Ralf Averbeck, Zhongbao Yin
submitted to PRD: <https://arxiv.org/abs/2404.17272>
- **Article: ALICE measures a rare charm baryon channel**
submitted to CERN Courier, to be published in <https://alice.cern/alice-cern-media>
- **Proceeding: Measurements of charm production in pp collisions with ALICE**
T.Cheng on behalf of ALICE Collaboration
SicPost Phys. Proc. 10.030 (2022)
- **Proceeding: Hadronisation of heavy quarks in small systems with ALICE at the LHC**
T.Cheng on behalf of ALICE Collaboration
submitted to Il Nuovo Cimento C, Colloquia and Communication in Physics
- **Proceeding: Charm production and hadronisation in pp and p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV at the LHC with ALICE**
T.Cheng on behalf of ALICE Collaboration
submitted to workshop organiser, to be published in Multidisciplinary Digital Publishing Institute (MDPI)
- **Proceeding: Measurement of branching-fraction ratio $\text{BR}(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e)/\text{BR}(\Omega_c^0 \rightarrow \Omega^- \pi^+)$ at ALICE**
T.Cheng on behalf of ALICE Collaboration
submitted to Nuclear and Particle Physics Proceedings



- **Talk: Ω_c^0 production in pp collisions at $\sqrt{s} = 13$ TeV with ALICE**
DPG conference March 2024, in Gießen, Germany
- **Talk: Measurement of branching-fraction ratio of $\text{BR}(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e)/\text{BR}(\Omega_c^0 \rightarrow \Omega^- \pi^+)$ at ALICE**
China LHC Physics workshop, November 2023, in Shanghai, China
- **Talk: Measurement of branching-fraction ratio of $\text{BR}(\Omega_c^0 \rightarrow \Omega^- e^+ \nu_e)/\text{BR}(\Omega_c^0 \rightarrow \Omega^- \pi^+)$ at ALICE**
FSP workshop. September 2023, in Lichtenfels, Germany
- **Talk: Hadronisation of heavy quarks in small systems with ALICE at the LHC**
Hadron 2023 conference, June 2023, in Genova, Italy
- **Talk: Ω_c^0 production in pp collisions at $\sqrt{s} = 13$ TeV with ALICE**
DPG conference March 2023, in Dresden, Germany
- **Online talk: Charm production and hadronisation in pp and p–Pb collisions at $\sqrt{s} = 5.02$ TeV at the LHC with ALICE**
Hot Quarks, October 2022, in Colorado, USA
- **Online talk: Measurements of charm production and hadronisation in pp collisions with ALICE**
The 50th International Symposium on Multiparticle Dynamics
- **Oneline talk: Ξ_c^0 and Ξ_c^+ baryon production in pp collisions with ALICE**
The 6th China LHC Physics workshop, November 2020, in China

- **Poster: Ω_c^0 production in pp collisions at $\sqrt{s} = 13$ TeV with ALICE**
LHCC poster session, November 2023, in Geneva, Switzerland
- **Poster: Ω_c^0 production in pp collisions at $\sqrt{s} = 13$ TeV with ALICE**
Quark Matter 2023, the XXXth International Conference on Ultra-relativistic Nucleus-Nucleus Collisions , September 2023, in Houston, USA
- **Poster: Ξ_c^0 production in pp collisions at $\sqrt{s} = 5.02$ TeV with ALICE**
Quark Matter 2019, the XXVIIIth International Conference on Ultra-relativistic Nucleus-Nucleus Collisions , November 2019, in Wuhan, China



致 谢

I would like to acknowledge all the people who, over the four years of my Ph.D., contributed to my development, not only in work but also in life. Without their support, I would not have been able to complete this journey.

First and foremost, I would like to thank my supervisor, **Zhongbao Yin**, who has supported all my activities since my master's and provided opportunities to attend international conferences. More than half of my Ph.D. life was spent abroad, and I am deeply grateful to **Silvia Masciocchi** for offering me the possibility to work with the wonderful ALICE@GSI group. Big thanks to my official co-supervisor, **Ralf Averbeck**, who has been incredibly nice and patient in answering my questions all the time. I am also grateful to **Andrea Dubla**, under whose guidance I have been working since 2019 on two analyses about charm-baryon production.

My two years in Darmstadt were beautiful experience, despite a tough year in 2022 when my analysis was not going well. Special thanks to Luuk for his valuable suggestions, discussions, and constant encouragement. Of course, the sports spirit cannot be forgotten! Fede, you are an amazing person with your unlimited energy and random/crazy ideas. You inspired my first marathon in Florence, and without you, my life in Darmstadt would have been much less colorful. JJ, you are a magician in solving tricky questions, and I cherish the times we spent running and cooking together. GG, you are the most patient person I have ever met. Thank you for introducing me to bouldering and climbing. Malu, I learned so much about Indian food and culture from you. With all of you, I have learned to balance life and work: only by enjoying life can one achieve better performance at work.

其次,感谢 ALICE 课题组的老师们。感谢周老师的鼓励和支持。感谢张晓明老师在课题分析上提出的宝贵建议。感谢毛亚显老师和裴骅老师在组会上给出的各种建议。感谢朱剑辉师兄的细心教导和讨论,从你这里,我学到了做事要面面俱到。也感谢粒子所里的所有授课老师们。

我要感谢:陈向娜,我的硕士博士七年室友,怀念我们一起做的早餐,以及一起吐槽过的剧;朱立强,感谢生活上的帮助,每次回国都去广埠屯地铁站接我;刘慧,可爱温柔的女生,从硕士到博士,我们总是‘擦肩而过’,你身上脚踏实地的科研精神一直激励我不断前行。虽然不是一个课题组,但是一直保持联系,斯图加特的大摆锤和 QQ 面是难忘的一天。卢鹏忠,感谢 GSI 这两年来在学术上的讨论,以及生活上的照顾。邢文静,非常优秀的理论师姐,在你身上看到了科研精神。怀念我们一起下午去健身房跑步的规律生活。也感谢在德国期间一起聚餐、嗨皮的



小伙伴们：刘慧、卢鹏忠、米珂、路坦、张定伟、喻佳俊、周博文、张留耀、张彪、周颖杰等。此外，还有国内其他小伙伴们对我科研和生活上的帮助：朱亚、崔捧瑶、郭文达、柳东海、方涛、崔凯、夏冰玉、李娜、张明宇、侯永珍、张震、易杰、何柳、徐冉、徐浪、李赏星、曹华琳、沈丽军、王瑞琴等。

特别感谢相识十几年的挚友，我的闺蜜郝佩佩，感谢这些年来陪伴和鼓励。虽然近几年我总在国外，但是仍旧倾听我的心事以及各种吐槽。感谢在北京、郑州的所有招待。在未来，我们要依旧保持联系，并常相见。

最后还要特别感谢我的家人，在我求学这一路上所有物质和精神上的支持，并尊重和支持我所有的决定。焉得萱草，言树之背，往后我会是你们的依靠。

行笔至此，我的学生时代即将杀青。往事堪堪亦澜澜，前路漫漫亦灿灿，愿依旧保持热忱之心，始终努力向上！

程甜甜

2024年5月20日于桂子山