Experimentelle Physik

Reconstruction of beauty jets in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with ALICE

Inaugural-Dissertation
zur Erlangung des Doktorgrades (Dr. rer. nat.)
der Naturwissenschaften im Fachbereich Physik
der Mathematisch-Naturwissenschaftlichen Fakultät
der Westfälischen Wilhelms-Universität Münster

vorgelegt von
KATHARINA DEMMICH
aus BOCHUM

– Münster, 2022 –

Contents

Contents

1. Introduction and motivation

It is fascinating that humanity these days has the possibility to uncover details about the composition and the fundamental interaction of matter down to length scales of smaller than 10^{-15} m – the approximate size of the proton [160]. This has been made possible by the construction of complex facilities that collide particles at high energies. In this sense, it was found in the 1970s based on data from the Stanford Linear Acceleration Center (SLAC) that the nucleons forming the atomic nucleus are composed of smaller objects called *quarks* [58, 61] which, under normal conditions, are bound into *hadrons* by the *strong interaction*. This discovery was groundbreaking and led to the formulation of the *Standard Model* as the theoretical foundation of particle physics [138].

Nowadays, proton-proton (pp) collisions at the so far highest possible energies are evoked using the Large Hadron Collider (LHC) at the European Organisation for Nuclear Research (CERN). The four major LHC experiments ALICE [8], ATLAS [2], CMS [72] and LHCb [44] collect data on the particles that participate in these collisions, each with an individual focus on a particular branch of particle physics. The analysis that is presented in this thesis is based on data from the ALICE experiment which is specialised on the investigation of the Quark-Gluon Plasma (QGP). This exotic state of matter in which gluons – the force carriers of the strong interaction – and quarks exist freely can be generated by relativistic collisions of heavy ions. The ALICE sub-detectors are thus designed for the operation in the dense environment of lead-lead (PbPb) collisions as the acronym **A L**arge **I**on **C**ollider **E**xperiment indicates. However, ALICE also pursues a dedicated program for pp collisions where it can reconstruct particles down to low momenta.

As of today, it is assumed that quarks exist in six *flavours* – up, down, strange, charm, top and beauty. After they are produced in high-energy collisions, quarks and gluons with high momenta generate sprays of particles called *jets*. In this thesis, a measurement of the cross section for beauty (b) jets, meaning jets initiated by a b quark, is presented for pp collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV using ALICE data from 2017. The motivations for this measurement shall be outlined in the following.

Due to their significant mass, the production of the *heavy flavours* charm, top and beauty in high-energy collisions is expected to be described by the theory of perturbative Quantum Chromodynamics (pQCD). Thorough testing of pQCD predictions by the confrontation with measurements on the heavy-flavour production is essential to assess the validity of the underlying theory. Predictions for the b-jet cross section in pp collisions by next-to-leading order (NLO) pQCD calculations in combination with parton shower algorithms based on Monte-Carlo (MC) techniques have been found to be compatible with previous measurements [3, 21, 77]. The measurement presented in this thesis leads the path to a subsequent confrontation of data and theory down to lower jet momenta in comparison to a respective available analysis by CMS [77]. The production of particles in high-energy collisions can also be estimated solely by the application of MC event generators. One generator that is widely used for the description of pp collisions is the leading-order (LO) generator PYTHIA [56]. In the past, it has been found that PYTHIA predictions show significant difficulties in describing measurements of the b-jet production [73, 77] which is largely driven by NLO processes [112]. The b-jet cross section presented in this thesis is compared to PYTHIA 8 predictions to provide further information with respect to the

modeling of heavy-flavour production.

The measurement of b jets is also interesting in the context of investigations of the QGP: In heavy-ion collisions, beauty quarks are generated before the emergence of the QGP and thus they experience the whole evolution of the hot medium. By measuring the properties of beauty observables, one can thus examine the properties of the QGP [123]. Thereby, respective measurements in pp collisions serve as a baseline for measurements in heavy-ion collisions as it is unlikely that a QGP is formed in small collision systems.

As demonstrated in the previous paragraphs, the applications of b-jet measurements are manifold. However, the possible measurement of the *dead-cone effect* for beauty quarks shall be highlighted in this context. The dead-cone effect [87, 88] denotes a prediction by pQCD that gluon emissions with small splitting angles are suppressed for heavy flavours in comparison to light flavours. Its first direct observation was proclaimed in 2019 by the ALICE experiment for gluon emission off charm quarks [15]. In the respective analysis, jets that contain a D^0 meson, and therefore have a large probability to originate from a charm quark, have been analysed. A decisive ingredient for the success of the applied techniques was the full reconstruction of the D^0 meson which made it possible to eliminate hadron-decay splittings that otherwise fill the dead cone. The investigation of the dead-cone effect for heavy flavours is particularly interesting as its precise measurement can provide information about the bare mass of the respective quark [15].

The measurement of the dead-cone effect for beauty quarks by uncovering the history of gluon emissions in b jets is theoretically possible. However, a full reconstruction of hadrons containing beauty quarks is not feasible with data from ALICE for previous data-taking periods [98]. Whether a direct measurement of the dead-cone effect for beauty quarks following the analysis strategy outlined above is still possible with the available ALICE data will be investigated in studies with the MC event generator PYTHIA as a part of this thesis.

Apart from motivations from the theory of particle physics, the objective of this thesis is to refine methods for the event-based identification of b jets – also referred to as b-jet *tagging* – that are applied by the ALICE collaboration. The special focus is thereby set on improving the performance at low transverse momenta. Previous analyses [3, 21, 73, 93] have shown, that the selection of jets that contain particle trajectories with a large displacement from the primary collision vertex is well suited for the identification of b jets. This is why, the measurement that is presented in this thesis is based on corresponding selection criteria.

This thesis is structured as follows: The first chapter provides background information reaching from basic principles of particle physics to more specific concepts with respect to the current understanding and description of hadron-hadron collisions. In addition, a short overview of the significance of b-jet measurements in the context of QGP studies is provided in this chapter. Chapter 3 adds background information on the experimental facilities and detection techniques. The ALICE experiment is introduced with all sub-detectors and the reconstruction of particle trajectories, as it is implemented for ALICE, is summarised. The most important observable that is utilised in this thesis is the signed transverse impact-parameter significance which is applied for the b-jet tagging. This observable is defined in Chapter 4 embedded in a review of previous b-jet measurements.

The individual analysis steps of the reconstruction of b jets in pp collisions at 13 TeV are detailed in Chapter 5. The results for the measurement of the b-jet cross section are discussed at the end of this chapter. Chapter 6 covers the simulation study with PYTHIA 8 which has been performed to investigate the possibility to expose the dead-cone for b jets in ALICE data. Projections of so-called *Lund Planes* [90] are performed to examine the impact of B-meson decay splittings on the signal region. The thesis concludes with a summary in Sect. 7.

2. Theoretical background

2.1. Basic concepts of particle physics

This section aims at providing a short introduction into the field of particle physics to create a baseline for the subsequent discussion. The following subsections are to a large extend based on [134, 153]; other sources will be cited, separately.

2.1.1. The Standard Model

For many years, the *Standard Model* (SM) of particle physics has been and is still today the foundation of our description of the existing particle species, their properties and their interactions. Nowadays it is assumed, that four fundamental interactions determine the interplay of the elementary particles: the electromagnetic interaction, the weak interaction, the strong interaction and gravity. The former three are described as parts of the SM while gravity, being almost forty orders of magnitude smaller than the strong force, is negligible for the majority of applications in particle physics. The significances of the remaining three interactions can be explained by looking at the structure of and the basic processes within an atom. The latter is formed by a compact nucleus built from protons (p) and neutrons (n) as well as electrons (e$^-$) that enclose the nucleus. The *electromagnetic interaction* acts on particles that carry electric charge and magnetic moment which is why the negatively charged electrons are confined to surround the positively charged nucleus. Mathematically, the electromagnetic (EM) interaction is described by Quantum Electrodynamics (QED). Within the nuclei of several isotopes, protons and neutrons can decay radioactively via β decays. This process is driven by the *weak interaction* that also determines the fusion processes in the sun. The protons and neutrons, which can be combined under the term nucleons, are not elementary but consist of smaller particles called quarks. These quarks are tied together by the *strong interaction*. In addition to this, the residual impact of the strong interaction also binds the protons and neutrons in the nucleus. The physics of the strong interaction is embedded into the theory of Quantum Chromodynamics (QCD).

Atoms consist of three species of elementary particles: the up (u) and down (d) quarks that form protons (uud) and neutrons (udd) and the electrons. However, further particle species exist in the SM and can be grouped into different categories, as it is shown in Fig. 2.1. The u and d quarks belong to a group of six quark species which is completed by the charm (c), strange (s), top (t) and beauty (b) quarks. Quarks are fermions, meaning particles with half-integer spin, that take part in all three interactions of the SM.
Apart from the quarks, there are six lepton species: electrons (e), muons (μ) and tauons (τ) together with the corresponding neutrinos ν_l, whereby $l = $ e, μ, τ. Leptons are fermions that participate in the weak interaction and, apart from the neutrinos, also in the EM interaction. Quarks and leptons can be arranged in three generations as it has been done in Fig. 2.1. The first generation is formed by the lightest quarks u and d, the electron and the electron neutrino. From the first to the third generation, the mass of the quarks is increasing and the up-type quarks carry an EM charge of $+\frac{2}{3}$ whereas down-type quarks carry a charge of $-\frac{1}{3}$. The beauty quark is of major interest for this thesis. It is the lighter quark of the third generation and has a mass of about $4.2\,\text{GeV}$ [160]. For every quark and lepton, there exists a respective antiparticle
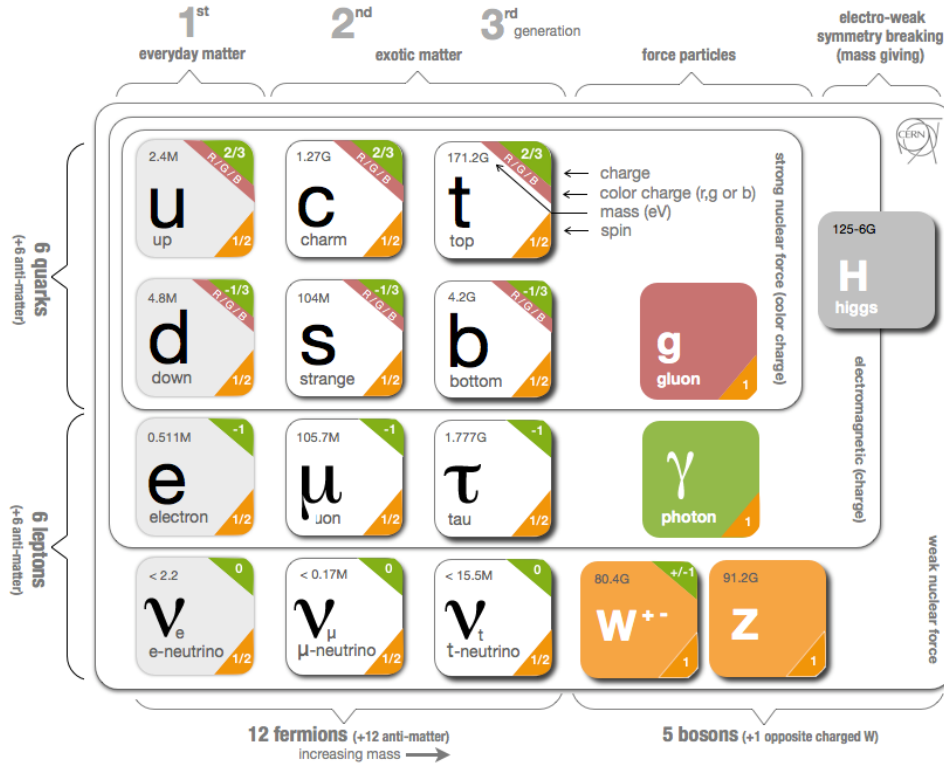
Figure 2.1. – Categorisation of the particles described by the Standard Model of particle physics [136, modified]. See text for more details.

with the same mass but opposite charge-like quantum numbers.

The last group of particles is formed by the gauge bosons which are spin-1 particles acting as force carriers of the three SM interactions. The massless photon represents the force carrier of the EM interaction although it does not carry EM charge itself. The massive W^+ and W^- bosons are electrically charged and transfer the charged current of the weak interaction. They are complemented by the massive Z^0 boson which transfers the weak neutral current between particles of the same electric charge. The force carriers of the strong interaction are the massless gluons. They carry colour charge, the charge of the strong interaction, which leads to several peculiarities of QCD with respect to QED that will be described in the next subsection.

Finally, there is the spin-0 Higgs boson which provides mass to the SM leptons, quarks and massive gauge bosons.

2.1.2. The strong interaction

As it was discussed before, the strong interaction couples to the so-called colour charge and is mediated by gluons – the gauge bosons of the strong interaction. Colour charge comes in three different colours, red, blue and green, as well as the three corresponding anticolours. There are eight types of gluons that each carry a colour and an anticolour. The fact that gluons are colour-charged allows them to self-interact and produce three-gluon or four-gluon vertices; something that is impossible for the photon in QED. One property of QCD that is believed to follow from this is the concept of *colour confinement*, which states that only colour-neutral objects exist freely in the QCD vacuum. Colour neutrality is achieved either by pairing a colour and an anticolour or by the combination of three different colours. Quarks carry only one (anti-)colour. This explains why they have so far only been observed as bound mesons (q$\bar{\text{q}}$ combinations) or
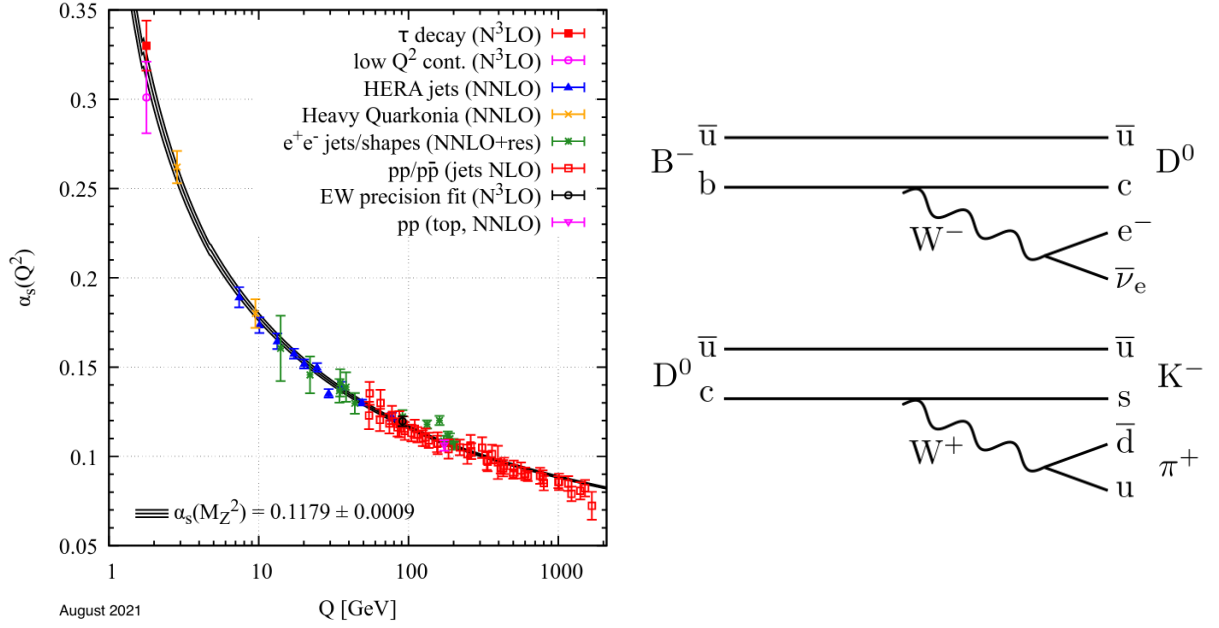
Figure 2.2. – **Left:** The dependence of the dimensionless strong coupling constant $\alpha_s(Q^2)$ on the virtuality Q^2 of the interaction [160]. **Right:** An example for the weak-decay cascade of the B^- meson via an intermediate D^0 meson into two leptons, a K^- and a π^+ meson.

baryons (qqq or $\overline{qqq}$ combinations) as well as more exotic compositions like tetra-, penta- or hexaquarks.

The strength of the interaction between two coloured objects is described by the dimensionless QCD coupling constant. At LO perturbation theory, this constant is given by [153]

$$\alpha_s(Q^2) = \frac{\alpha(\mu^2)}{1 - \alpha_s(\mu^2)\frac{1}{3\pi}\ln\left(\frac{Q^2}{\mu^2}\right)}. \tag{2.1.1}$$

Eqn. 2.1.1 needs the input from a known value of $\alpha_s(\mu)$ at some scale μ and depends via the virtuality $Q^2 = -q^2$ on the momentum transfer q of both particles. Particles participating in the strong interaction experience a smaller coupling for large Q^2 and a larger coupling for smaller Q^2, as it can be seen from the distribution of α_s in Fig. 2.2. Phenomenologically, this behaviour can be assigned to the effect of *anti-screening*: quantum fluctuations lead to the appearance of additional gluons that screen bare colour charges. The larger the momentum with which one colour charge perceives another, the more gluons it resolves on the way. This opens the possibility for colour charges to exist quasi-freely at large Q^2, an effect called *asymptotic freedom*. In this case, particle production and interaction via the strong interaction can be described by perturbation theory, i.e. by pQCD, as expansions in α_s. In contrast, phenomenological models need to be applied at low Q^2.

In the following paragraph, the theoretical foundation of the strong interaction by the theory of QCD shall be introduced, as it is described in [147, 153]. To do so, it is expedient to start from the Dirac equation

$$\left(i\gamma^\mu\delta_\mu - m\right)\psi = 0, \tag{2.1.2}$$

which describes relativistic spin$-\frac{1}{2}$ particles – here represented by their wave function ψ. QCD is a quantum field theory that is based on the assumption that "physics", and thus the Dirac equation, is invariant under SU(3) local gauge transformation in the form of

$$\psi(x) \to \psi'(x) = \exp\left(ig_s\theta^a(x)T^a\right)\psi(x). \tag{2.1.3}$$

Thereby, T^a with $a \in [1, \ldots, 8]$ are the eight 3×3 generators of the SU(3) symmetry group that can be represented by the Gell-Mann matrices λ^a (see [153] for definition) via

$$T^a = \frac{1}{2}\lambda^a \tag{2.1.4}$$

and g_s is the QCD coupling constant that is related to α_s by $g_s^2 = 4\pi\alpha_s$.

The requirement of local gauge transformation necessitates the introduction of eight new fields G_μ^a which can be identified as the eight gluons fields. To meet the requirements of local gauge invariance, these fields need to transform as

$$G_\mu^c \to G_\mu^{c'} = G_\mu^c - \delta_\mu\theta_c - g_s f_{abc}\theta_a G_\mu^b. \tag{2.1.5}$$

The Dirac equation from Eqn. 2.1.2 is extended by a term that describes the interaction of G_μ^a with the charged-particle wave function according to

$$i\gamma^\mu \left[\delta_\mu + ig_s G_\mu^a T^a\right]\psi - m\,\psi = 0, \tag{2.1.6}$$

In this context, three new degrees of freedom of the wave function ψ are introduced which correspond to the three colour states red, blue and green and can be represented by three-component orthogonal vectors. The gauge transformation of Eqn. 2.1.3 describes a rotation of these colour states in colour space that is carried out by the eight gluons of the strong interaction, as represented by the matrices λ^a. The last term in Eqn. 2.1.5 is the mathematical origin of the gluon self-interaction which makes QCD a non-abelian gauge theory. The term arises from the fact that the gluon fields do not commute but the commutator is given by the SU(3) *structure functions* f^{abc} with $a, b, c \in [1, \ldots, 8]$ via

$$[\lambda^a, \lambda^b] = 2if^{abc}\lambda_c. \tag{2.1.7}$$

2.1.3. Beauty-hadron production and decays

The total production cross section for $b\overline{b}$ pairs in $13\,\mathrm{TeV}$ pp collisions can be averaged from measurements by LHCb [6] and ALICE to about $502\,\mu\mathrm{b}$ [23]. In the following, the fragmentation of these quarks into beauty hadrons and the subsequent decay of the latter into detectable final-state particles will be discussed.

Beauty quarks fragment into different beauty-hadron species, of which a selection is listed in Tab. 2.1 together with their quark content. Thereby, the fragmentation process depends on the production mechanisms and the transverse momentum of the initiating beauty quark. Only few measurements of these fragmentation fractions are available for high-energy collisions from LHC experiments. The LHCb experiment has measured the ratio of the fragmentation fraction into $\overline{B}_s^0$ mesons and Λ_b^0 baryons over the one into the non-strange beauty mesons $\overline{B}^0$ and B^- with about $12\,\%$ and $26\,\%$ for $13\,\mathrm{TeV}$ pp collisions [7]. More detailed measurements are available for $p\overline{p}$ and e^+e^- collisions [47, 160].

The B meson ground states are pseudoscalar mesons ($J^P = 0^-$) with a mass of about $5\,\mathrm{GeV}/c^2$ which cannot decay via the flavour-preserving strong or EM interaction, but merely via the flavour-changing charged current of the weak interaction. Thereby, the coupling of the beauty quark to the other quark flavours are described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix which relates the weak eigenstates ($q' \in [\mathrm{d}', \mathrm{s}', \mathrm{b}']$) and the mass eigenstates ($q \in [\mathrm{d}, \mathrm{s}, \mathrm{b}]$) of the quarks as

$$\begin{pmatrix} \mathrm{d}' \\ \mathrm{s}' \\ \mathrm{b}' \end{pmatrix} = \begin{pmatrix} V_\mathrm{ud} & V_\mathrm{us} & V_\mathrm{ub} \\ V_\mathrm{cd} & V_\mathrm{cs} & V_\mathrm{cb} \\ V_\mathrm{td} & V_\mathrm{ts} & V_\mathrm{tb} \end{pmatrix} \begin{pmatrix} \mathrm{d} \\ \mathrm{s} \\ \mathrm{b} \end{pmatrix}. \tag{2.1.8}$$

Table 2.1. – The quark content and the world averages of the lifetimes of several beauty-hadron species taken from [160].

Meson	Quark Content	Lifetime [ps]
B^+	$u\overline{b}$	1.638 ± 0.004
B^0	$d\overline{b}$	1.519 ± 0.004
B_s^0	$s\overline{b}$	1.515 ± 0.004
B_c^+	$c\overline{b}$	0.510 ± 0.009
Λ_b^0	udb	1.471 ± 0.009
Ξ_b^-	dsb	1.572 ± 0.040
Ξ_b^0	usb	1.480 ± 0.030
Ω_b^-	ssb	$1.64^{+0.18}_{-0.17}$
$\Upsilon(1S)$	$b\overline{b}$	$(1.22 \pm 0.03)\,10^{-8}$

Since the CKM matrix is close to a unity matrix and the decay of the beauty quarks is a generation-changing process, beauty hadrons, apart from bottomonia, exhibit comparatively long lifetimes with $c\tau \approx 500\,\mu\text{m}$ [160]. Their decay can be described by the *spectator model* as the decay of the beauty quark with the dominant decay mode b→cW^{*-}. The virtual W^{*-} boson further decays into a pair of leptons $l\overline{\nu}$ or into a $q\overline{q}$ pair. This decay chain leads to the large branching fractions of B mesons into D mesons (e.g. BR(B$^-\to D^0 X$)=$(79 \pm 4)\,\%$ [160]) which themselves decay with a significant branching fraction via semileptonic decays and decays into kaons. An example for the weak-decay cascade of a B$^-$ meson is shown in Fig. 2.2. Bottomonia, like the Υ states, are mesons that are formed by b$\overline{b}$ pairs. They can decay via the electromagnetic and the strong interaction, which is why they exhibit lifetimes that are smaller by several orders of magnitude than those of B mesons. The lifetime of the dominant beauty-hadron species are summarised in Tab. 2.1 [31, 153, 160].

2.2. Hadron-hadron collisions

The collisions of hadrons at collider facilities like the LHC at CERN offer the possibility to study the strong interaction in detail and to test corresponding predictions by the Standard Model. This chapter discusses basic knowledge on the physics of inelastic hadron-hadron collisions whereby it concentrates on collisions with large initial momentum transfer in line with the focus of the analysis in this thesis.

The definition of basic kinematic observables are not elaborated on in this thesis. In this context, it is referred to dedicated text books.

2.2.1. Categorisation of basic processes

For the majority of pp collisions at the LHC, the momentum transfer of the initial interaction is small. In these cases, either one or both of the protons get exited and decay, which is denoted as diffractive scattering [64], or the protons may interact elastically. In this thesis, however, the focus is set on the fraction of collisions with large initial momentum transfer for which both protons break up into partons that are assumed to be quasi-free. A sequence of processes is initiated over a range of different energy scales. Thereby, *soft processes* with small momentum transfer $Q^2 = -q^2 \lesssim 1\,\text{GeV}$ can not be described perturbatively in contrast to processes with significantly larger momentum transfer which are referred to as *hard processes* [160]. The different kinds of processes that take place in pp collisions with large initial momentum transfer are shown schematically in Fig. 2.3 and can be categorised as follows:

- **Initial state radiation (ISR):** The partons in each incoming hadron may already radiate gluons or photons before the actual hadron-hadron collision and thus produce ISR that can affect the later collision history.

- **Hard parton interactions:** The partons from both hadrons interact via hard scattering or annihilation processes. These processes with large momentum transfer determine the basic properties of the collision event and act as the seeds for the production of bundles of high-momentum particles called *jets*. Since hadrons are composed of several partons, multiple hard parton interactions can take place in one hadron-hadron collision. Often, only the interaction with the largest Q^2 is referred to as hard(est) process whereas interactions with lower Q^2 are combined under the name *multi-parton interactions* (MPIs).

- **Final State Radiation (FSR):** Particles that originate from the hard interactions can generate new partons of lower momentum via gluon radiation and the splitting of gluons into quark pairs.

- **Hadronisation:** As soon as the partons' energies drop below about $1\,\text{GeV}$, quarks and gluons are no longer asymptotically free and they combine to form hadrons.

- **Particle decays:** Bound states which are generated directly via hadronisation processes are called *primary* hadrons and they can further decay into lighter hadrons. Thereby, all hadrons that are created via electromagnetic and strong decays shall also be referred to as primary hadrons in this thesis whereas hadrons originating from weak decays will be referred to as *secondary* hadrons. For more information on the ALICE definition of primary particles see [151].

Any activity (ISR, MPIs ...) which is not directly related to the hardest process is, in the context of the analysis of hard objects, referred to as the *underlying event*. As the underlying event cannot be described perturbatively, it needs to be subtracted before comparing measurements to pQCD calculations.

2.2.2. Model description of the hadron-production cross section

An important basis for the mathematical description of particle production in hadron-hadron collisions is the possibility to separate low-energy (= large-distance) from large-energy (= small-distance) processes via the factorisation of the production cross section according to [64, 76, 135]

$$\mathrm{d}\sigma^h = \sum_{a,b,n} f_a(x_a, Q^2) \otimes f_b(x_b, Q^2) \otimes \mathrm{d}\sigma_{ab \to n} \otimes D_{n/h}(z, Q^2). \tag{2.2.1}$$

The individual factors denote the Parton Distribution Functions (PDFs) $f_a(x_a, Q^2)$ and $f_b(x_b, Q^2)$ of the initial hadrons, the parton-level production cross section $\sigma_{ab \to n}$ for the description of the hard interaction and the fragmentation function $D_{n/h}(z, Q^2)$ that makes the partonic final state n evolve towards a final-state hadron h. Their meaning, especially with respect to heavy-flavour production, shall be described in detail in the next paragraphs.

Parton distribution functions   The PDFs $f_i(x_i, Q^2)$ describe the abundance of partons within the nucleons depending on the fractional momentum x_i of the parent hadron. They cannot be calculated based on perturbative QCD as the coupling constant is too large at the energy scales of interest. Instead, they are determined experimentally from measurements of deep inelastic lepton-nucleon scattering and hard scatterings in nucleon-nucleon collisions [160]. The proton PDFs for the different parton species, as they have been determined by the NNPDF collaboration [51], are shown exemplarily in Fig. 2.4 for two energy scales. It becomes apparent that at low x the gluon and sea quark PDFs are dominating whereas at larger x the PDFs of the constituent quarks are most significant.
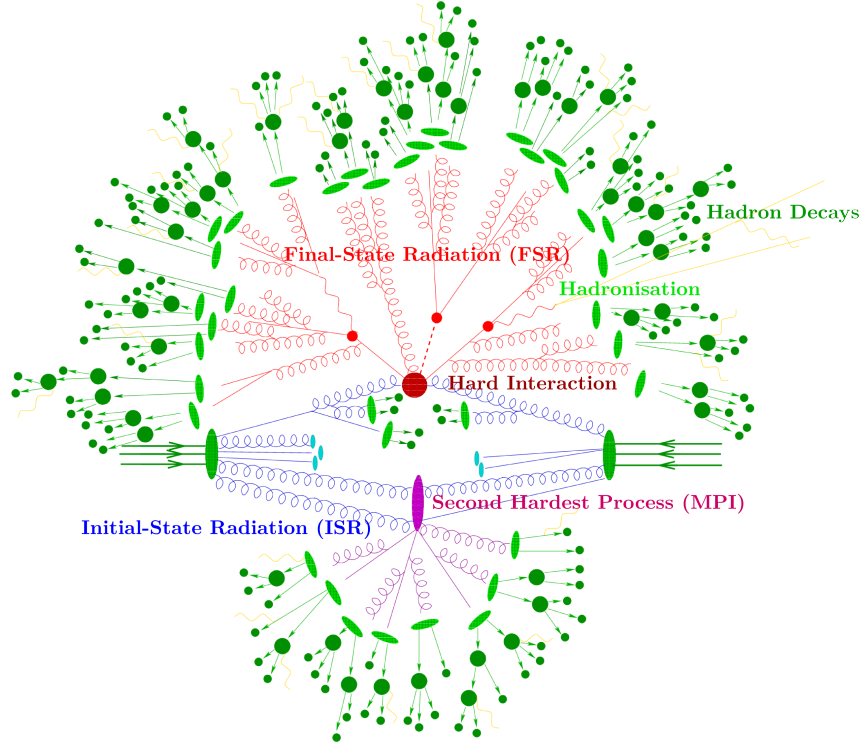
Figure 2.3. – Schematic representation of the different processes that take place during and after a hadron-hadron collisions [105, modified]. See text for more details.
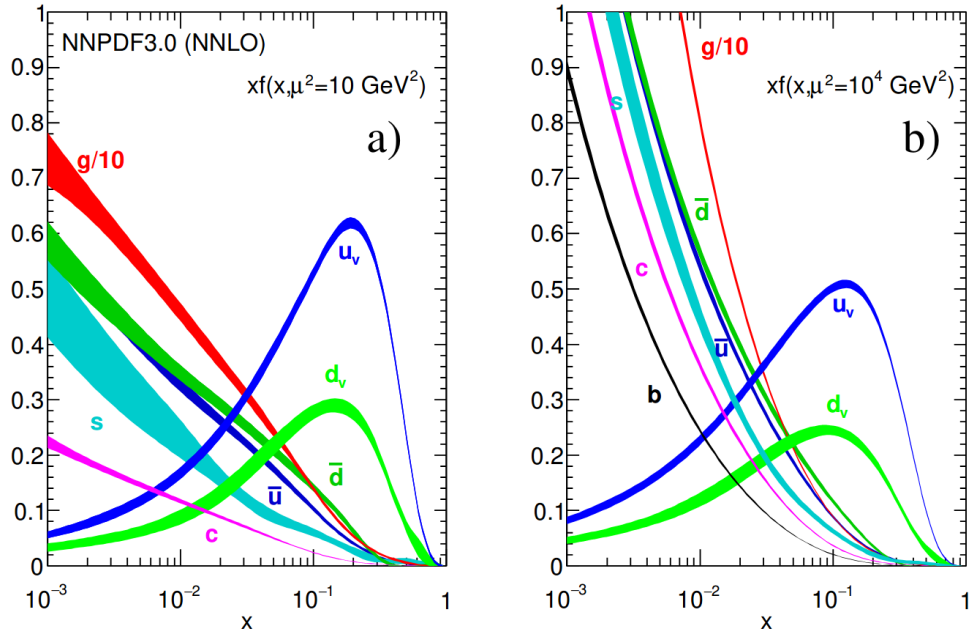
Figure 2.4. – The parton densities xf within the proton determined by the NNPDF collaboration for two energy scales μ **a)** $\mu^2 = 10\,\mathrm{GeV}^2$ and **b)** $\mu^2 = 10^4\,\mathrm{GeV}^2$ [160]. These results have been determined based on data from HERA as well as LHC experiments [51].
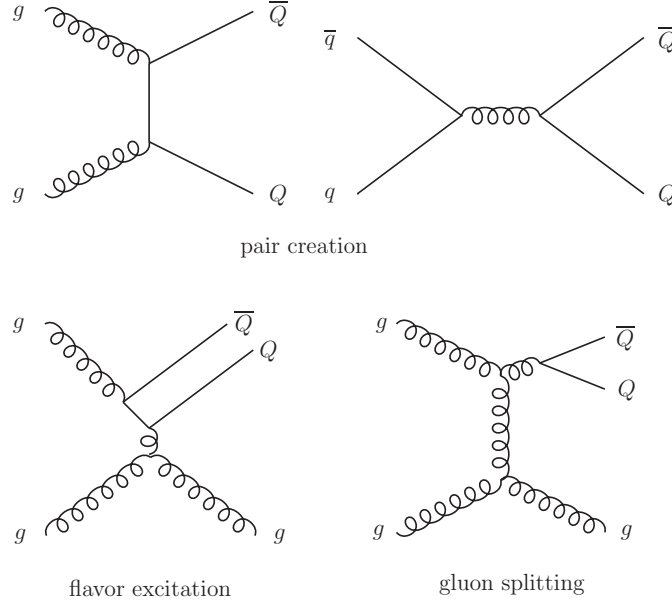
Figure 2.5. – The three Feynman diagrams that are most important for the production of heavy flavours Q from light flavours q and gluons g: pair creation via the annihilation of light flavours or gluons, flavour excitation in the initial state and the splitting of gluons in heavy quark-antiquark pairs.

Parton-level production cross section The transition amplitude of a particular parton interaction is described by Lorentz-invariant matrix elements. In particular, the cross section $\sigma_{ab \to n}$ for the transformation of the parent partons a and b into the partonic final state n is given by the sum of all matrix elements that lead to a final state $n + X$, integrated over the available phase space. Due to the large momentum transfer in hadron-hadron collisions, the corresponding matrix elements can be calculated perturbatively [64, 153].

Heavy flavours can be created at all orders of perturbation theory. Three types of Feynman diagrams are, however, of particular importance in this context: pair creation (LO) via gluon or quark fusion, flavour excitation (NLO) and gluon splitting (NLO) (see Fig. 2.5). It has been found that at low centre-of-mass energies, the production of beauty quarks primarily takes place via pair creation [29], while at LHC energies the contribution from NLO processes are dominating [112].

**Fragmentation function** The fragmentation function $D_{n/h}(z, Q^2)$ describes the evolution of the partonic final state n up to the formation of a hadron h which carries a fractional momentum z of the initiating parton. An analytic description of this process is extremely challenging due to the large phase space and the huge multiplicity of intermediate multi-particle states. Nevertheless, a numeric description is possible with MC event generators and can be divided into two parts: the *parton shower* that connects to the particles outgoing from the hard process and evolves them perturbatively via a Markov chain sequence of splittings and the hadronisation mechanism that performs the non-perturbative process of hadron formation via phenomenological models [64, 135].

With respect to heavy-flavour production, partons with large mass can be generated via high-energy gluon splittings in the parton shower. The production of heavy flavours during hadronisation is negligible due to the low energy scales involved.

2.2.3. Particle jets

In high-energy particle collisions, the quarks or gluons that are outgoing from the hard process produce collimated bundles of new particles which are called jets. From a theorist's perspective,

a jet at LO is defined as the whole bunch of particles that is produced by a single outgoing particle. In this idealised case, the jet carries information on the properties of the hard interaction in general, the initiating parton and its fragmentation. Jet-finding algorithms approximately recover the ideal object by clustering particles in a certain angular volume or with similar momentum. Nevertheless, jets objects differ significantly if they are obtained using different clustering algorithms: the jet is defined by the algorithm and the algorithm's settings which were used to select it.

A critical property of a jet algorithm is referred to as *soft and collinear safety* (IRC safe) which means that the objects the algorithm selects are stable with respect to soft and collinear emissions. If a particle is replaced by a collection of soft or collinear particles with the same total momentum then the jet's properties should not change [64].

Available jet clustering algorithms can be divided into two major groups: *cone algorithms* which gather all particles that can be found into a cone with certain opening angle and *sequential recombination algorithms* that cluster objects based on their momentum and angular difference. Most of the cone algorithms are not IRC safe and are thus rarely used nowadays. The sequential recombination algorithms all follow a similar basic structure which can be sketched as:

Sequential Recombination Algorithm [48, 68]

---

while particles in event **do**
    for particles i, j in event **do**
        calculate distance
$$d_{ij} = \min(p_{\mathrm{T},i}^{2p}, p_{\mathrm{T},j}^{2p})\frac{\Delta_{ij}^2}{R^2} \text{ with } \Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\varphi_i - \varphi_j)^2$$
        and distance to beam in momentum space
$$d_{iB} = \frac{1}{2p_{\mathrm{T},i}^2}$$
    if any $d_{ij} < d_{iB}$ **then**
        combine particles i and j within one jet
    else
        particle i is final jet candidate
        remove i from list of particles

Thereby, $p_{\mathrm{T},i}, \varphi_i$ and η_i are the transverse momentum, azimuthal angle and pseudorapidity of particle i. The *jet radius R* restricts the opening angle of the cone that encompasses all constituents as it suppresses the clustering of particles with large angular distance. The exponent p regulates the importances of the particles' momenta and angle with respect to the clustering decision and defines the following IRC safe algorithms [48, 68, 140]

- **Cambridge-Aachen algorithm** ($p = 0$): Clustering is only based on the relative angle Δ_{ij} of particles i and j: particles closest in space are clustered first. Since soft gluon emissions are approximately ordered according to their emission angle [15, 87], the algorithm performs well with respect to capturing the history of particle splittings.

- $k_{\mathbf{T}}$ **algorithm** ($p = 1$): Soft particles that are close in space are clustered first. Similar to the Cambridge-Aachen algorithm, this ordering approximately reconstructs the splitting history. The fact that soft splittings are clustered early in the process makes the algorithm well suited for resolving subjets.

- **anti$-k_{\mathbf{T}}$ algorithm** ($p = -1$): Hard particles that are close in space are clustered first. The resulting jets are resilient with respect to background from the underlying event or overlapping pp collisions and thus the algorithm performs well with respect to identification of hard jets. In contrast to the Cambridge-Aachen and the k_{T} algorithm, the anti$-k_{\mathrm{T}}$ algorithm is unsuited for the investigation of jet substructure.
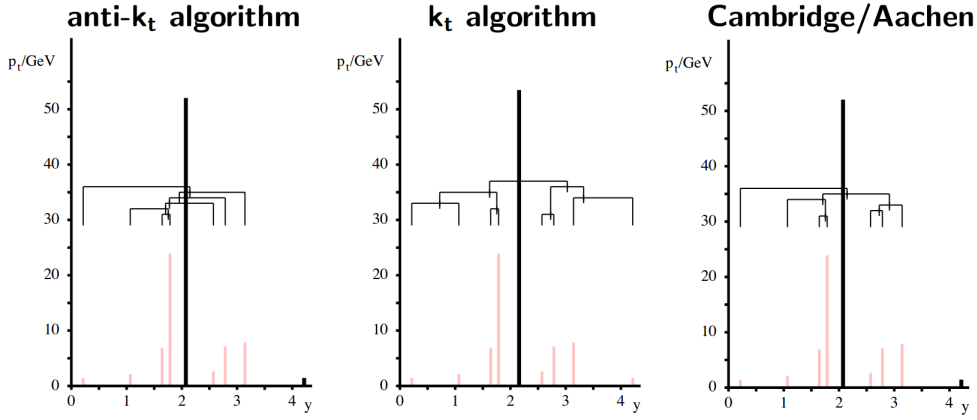
Figure 2.6. – The performance of the anti-k_T, the k_T and the Cambridge-Aachen algorithm for an exemplary event of particles with different transverse momenta p_T and different angular distances y. The anti-k_T algorithm combines soft wide-angle gluon emissions late in the clustering process: one hard jet is resolved that is resilient with respect to soft emissions. On the contrary, the k_T and the Cambridge-Aachen algorithm resolve two subjets which are combined to a single jet at the end of the clustering process [140].
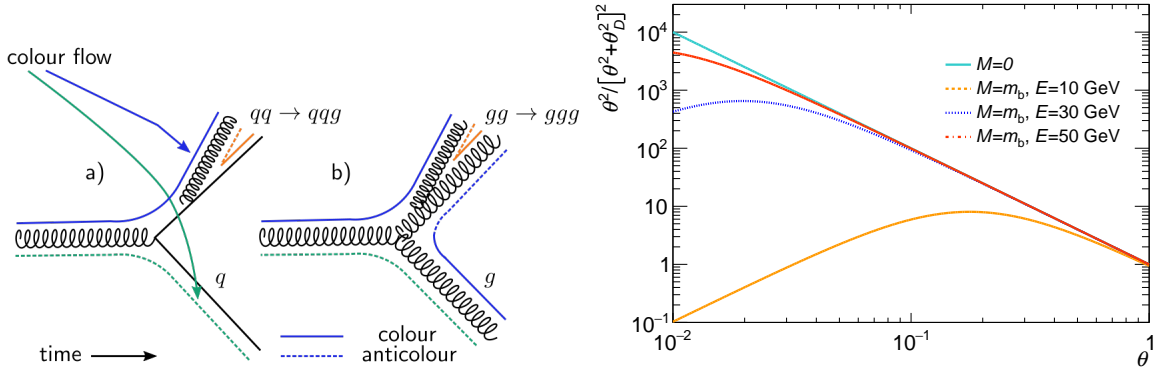
Figure 2.7. – **Left:** Two exemplary colour flow scenarios for the production of hard particles and a subsequent gluon emission a) generation of two quarks q with a subsequent $q \to qg$ emission, b) generation of two gluons g with a subsequent $g \to gg$ splitting. **Right:** The term from the definition of the probability for gluon radiation that describes the angular dependence, is shown for gluon radiation off massless particles ($M = 0$) and off beauty quarks ($M = m_\mathrm{b}$) with energy E of the radiator.

The characteristics of the listed jet algorithms can be illustrated based on their performance on an exemplary selection of particles with different momenta and angular distances as it can be seen in Fig. 2.6.

There are different *recombination schemes* which specify how the jet total four-momentum is calculated from the properties of the jet constituents. The E-scheme adds the four vectors of the particles that are combined and is widely used in particle physics. For more information on recombination schemes see [69].

2.2.4. Gluon radiation in the QCD vacuum

The probability for gluon radiation off quarks and gluons in the parton shower exhibits two major dependencies which are predicted by QCD: a dependence on the particle's colour factor and its mass. These dependencies will now be discussed in more detail, as the investigation of the mass dependence motivates the simulation study presented in Sect. 6.

Dependence on the radiator's colour factor To demonstrate the influence of the radiator's colour charge on the probability for gluon radiation, the cross section σ_{ppg} for the production

$X \to p_1 p_2 g$ of two hard particles p_1 and p_2 and a soft gluon from an arbitrary initial state X will be investigated. To start with, it will be assumed that the mass of the radiator is negligible; the mass dependence will be discussed later in this section.

For the calculation of σ_{ppg}, it needs to be taken into account that strongly interacting particles can occur in different colour charges. In general, several colour combinations or *colour flows* are allowed by colour conservation for each matrix element. Two colour-flow examples for a $X \to p_1 p_2 g$ process, one for the radiation off a quark ($qq \to qqg$) and one for the radiation off a gluon ($gg \to ggg$), are shown in Fig. 2.7. One can recognise from the figure, that since gluons carry two colour charges, the variety of colour flow scenarios is larger for the latter than for the former.

The cross section σ_{ppg} is given by the sum over all colour flow scenarios for the corresponding matrix element M_{ppg} averaged over the number of possible incoming colours N_c^{in}

$$\sigma_{ppg} \propto \frac{1}{N_c^{\text{in}}} \sum_{\text{colours}} |M_{ppg}|^2. \tag{2.2.2}$$

The colour information is encoded in the vertex factors that enter the matrix element calculation and can be defined based on the Gell-Mann matrices Λ^a. For each qqg vertex, a factor

$$- g_s \frac{i}{2} \Lambda_{ji}^a \gamma^\mu \tag{2.2.3}$$

with the colours $i, j \in [1, \ldots, 3]$ needs to be considered and for each ggg vertex a factor

$$- g_s f^{abc} V(p_1, p_2, p_3) \tag{2.2.4}$$

needs to be assigned. Hereby, γ^μ are the Dirac γ matrices and $V(p_1, p_2, p_3)$ is a function that depends on the gluon momenta p_i but does not contain any colour information. By squaring the matrix element according to Eqn. 2.2.2, the following terms arise

$$\sum_{c,d} f^{acd} f^{bcd} = C_A \delta^{ab} \quad \text{and} \quad \frac{1}{4} \sum_a \Lambda_{ij}^a \Lambda_{jk}^a = C_F \delta_{ik}. \tag{2.2.5}$$

The operator $C_A = 3$ thereby embodies the colour factor for the $g \to gg$ splitting while the operator $C_F = \frac{4}{3}$ describes the gluon radiation $q \to qg$ [147].

For the emission of a soft and collinear gluon, σ_{ppg} can be formulated in the *infrared and collinear limit* which is utilised in many MC event generators. In this limit, σ_{ppg} is calculated as the product of the cross section σ_{pp} for the hard production of p_1 and p_2 and the probability $\mathrm{d}P_{\text{LF}}(\theta, \omega)$ for a subsequent gluon emission with angle θ and energy ω [87, 139, 160]

$$\lim_{\omega \to 0, \theta \to 0} \mathrm{d}\sigma_{ppg} = \sigma_{pp} \underbrace{\frac{\alpha_s C_i}{\pi} \frac{\mathrm{d}\theta^2}{\theta^2} \frac{\mathrm{d}\omega}{\omega}}_{\mathrm{d}P_{\text{LF}}(\theta, \omega)}. \tag{2.2.6}$$

In Eqn. 2.2.6, the dependence for the radiation probability on the colour factor C_i with $i = A, F$ becomes obvious. On this basis, it can be concluded that gluons radiate more than quarks which is due to the larger variety of possible colour flow scenarios of these types of splittings [64, 147, 160].

Dependence on the radiator's mass So far, it has been assumed that the mass of the radiator is negligible and thus Eqn. 2.2.6 is an approximation for small masses. It can be shown using classical field theory that the probability $\mathrm{d}P_{\text{HF}}(\theta, \omega)$ for gluon radiation off a massive particle

with mass M and energy E exhibits a different angular dependence than $\mathrm{d}P_{\mathrm{LF}}(\theta, \omega)$ namely [87, 88]

$$\mathrm{d}P_{\mathrm{HF}}(\theta, \omega) = \frac{\alpha_s C_F}{\pi} \frac{\theta^2 \mathrm{d}\theta^2}{\left[\theta^2 + \theta_D^2\right]^2} \frac{\mathrm{d}\omega}{\omega} \quad \text{and} \quad \theta_D = \frac{M}{E}. \tag{2.2.7}$$

For small masses, the contribution from θ_D approaches zero and the angular dependence provided in Eqn. 2.2.6 is reproduced

$$\lim_{M \to 0} \underbrace{\frac{\theta^2 \mathrm{d}\theta^2}{\left[\theta^2 + \theta_D^2\right]^2}}_{\alpha_{\mathrm{HF}}(\theta)} = \underbrace{\frac{\mathrm{d}\theta^2}{\theta^2}}_{\alpha_{\mathrm{LF}}(\theta)} \tag{2.2.8}$$

From Eqn. 2.2.7, it can be read that gluon radiation is suppressed in a so-called *dead cone* for which the emission angle is smaller than the characteristic angle θ_D. At large angles, the probability for radiation off light and heavy flavours is universal [87, 88]. This behaviour is demonstrated by Fig. 2.7, where α_{HF} is shown for $M = m_b = 4.2\,\mathrm{GeV}/c^2$ [160] and three radiator energies $E = 10, 30$ and $50\,\mathrm{GeV}$. With increasing energy, the maximum of the radiation pattern is shifted towards smaller angles and the suppression with respect to the distribution for α_{LF} decreases. At large angles, all distributions approach a common curve.

The dead-cone effect has consequences which are defined in [88] and which enable to indirectly uncover it in measurements. First of all, the suppression of gluon radiation leads to a shift of the relative energy fraction that is carried by a heavy flavour to larger values compared to light flavours. This has been confirmed by measurements of the fractional beam energy that is carried by B hadrons in e^+e^- collisions, e.g. see [30]. Moreover, heavy-flavour events produce less light-flavour hadrons in addition to direct decay products of the heavy-flavour mesons. An analysis of the difference between the average charged particle multiplicity in $e^+e^- \to b\bar{b}$ and $e^+e^- \to q\bar{q}$ events ($q = u, d, s$) verified this behaviour [14].

The first direct observation of the dead-cone effect was announced by ALICE in 2019 for a measurement of D^0 jets in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$. This measurement constituted important progress as apart from directly addressing a fundamental prediction by QCD, the closer investigation of the mass-dependence of the dead-cone effect in principle allows for the direct measurement of the bare mass of the charm quark. The analysis applied different jet clustering algorithms to disentangle individual particle splittings in the charm fragmentation and access the corresponding splitting angles. A decisive ingredient for the success of the technique was the full reconstruction of the D^0 meson which made it possible to separate hadron decay splittings that fill the dead cone [15, 159].

2.2.5. The primary Lund Plane

With the help of the Lund Plane, it is possible to visualise the average density of primary particle splittings

$$\rho(\theta, k_{\mathrm{T}}) = \frac{1}{N_{\mathrm{Jets}}} \frac{\mathrm{d}^2 n}{\mathrm{d}\ln(1/\theta) \mathrm{d}\ln(k_{\mathrm{T}})} \tag{2.2.9}$$

in the history of jet candidates in dependence of the splitting angle θ and the transverse momentum k_{T} of the splitting as proposed in [90]. Here, *primary splittings* are gluon emissions off the initiating parton whereby splittings of higher order correspond to splittings off former gluon emissions as it is schematically shown in Fig. 2.8. The splitting angle θ is given by

$$\theta = \sqrt{(\eta'_{\mathrm{rad}} - \eta_{\mathrm{em}})^2 + (\varphi'_{\mathrm{rad}} - \varphi_{\mathrm{em}})^2} \tag{2.2.10}$$
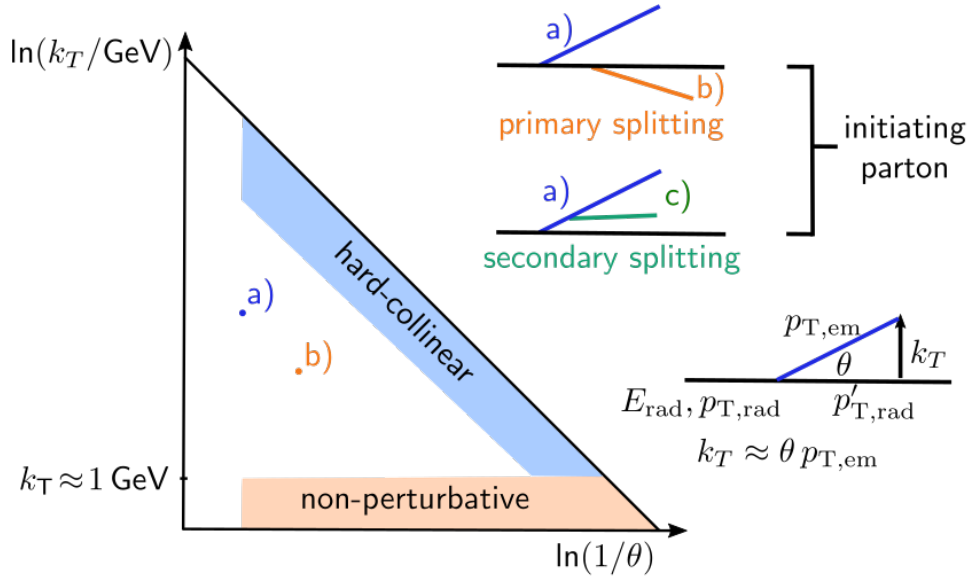
Figure 2.8. – Scheme of the primary Lund Plane which monitors the average density of primary gluon emissions in dependence of the emission angle θ and the transverse momentum k_T with respect to the radiator. The radiating parton carries the energy E_{rad} and the transverse momentum $p_{\text{T,rad}}$ before and $p'_{\text{T,rad}}$ after the emission. Primary emissions are defined as splittings of the initiating parton whereby secondary emissions represent emissions off a previous gluon emission. This figure is inspired from figures in [90].

with η_{em} and η'_{rad} being the pseudorapidity and φ_{em} and φ'_{rad} the azimuthal angle of the emitted gluon and the radiator after the emission. The transverse momentum of the splitting is defined as

$$k_{\text{T}} = \theta \, p_{\text{T,em}} \tag{2.2.11}$$

in the small angle approximation with the transverse momentum $p_{\text{T,em}}$ of the emitted gluon with respect to the beam. By measuring the Lund Plane for small energies E_{rad} of the radiator the dead-cone effect for heavy flavours can be uncovered [82].

As it was said in the previous section, the anti-k_{T} algorithm performs best with respect to jet identification but worse if it comes to substructure measurements. Therefore, to fill the Lund Plane, the anti-k_{T} algorithm is used for the jet finding but the identified jets are reclustered with the Cambridge-Aachen algorithm. In the subsequent declustering of the splitting history, the jet object j is iteratively split into subjets and for every splitting the transverse momentum k_{T} and the angle θ are filled in the Lund Plane. One always follows the branch with the largest momentum to assure the selection of primary splittings:

Iterative Declustering [82, 90]

while jet j can be split into subjets j_1, j_2 **do**
  if $p_{\text{T,1}} < p_{\text{T,2}}$ **then** swap j_1 and j_2
  treat j_1 as radiator, i.e. $p'_{\text{T,rad}} = p_{\text{T,1}} \ldots$
  treat j_2 as emitter, i.e. $p_{\text{T,em}} = p_{\text{T,2}} \ldots$
  calculate splitting properties (θ, k_{T}, $E_{\text{rad}} = E_1 + E_2, \ldots$)
  $j = j_1$

Using this algorithm, gluon splittings into q$\bar{\text{q}}$ pairs are misinterpreted as gluon emissions. For the declustering of beauty jets with regard to the investigation of the dead-cone effect, it would be preferable to follow the branch containing the beauty quark in the declustering process.

However experimentally, it is impossible to identify the corresponding branch and it is also not necessary: it has been found in [82] that the hardest branch matches the branch that contains the b quark in 99.999% of all cases.

One remarkable property of the primary Lund Plane is the possibility to separate the contribution of different processes to the splitting history via selecting or de-selecting particular regions of the diagram, as it is shown schematically in Fig. 2.8. The dead-cone effect can be investigated in the *hard-collinear* regime close to the confining diagonal at small θ and large k_T. With a single cut on k_T, the contribution from non-perturbative effects can be rejected which is dominating the region around and below $k_T \approx 1\,\text{GeV/c}$. Initial-state radiation and the underlying event fills the Lund Plane at large $\theta \approx 1$ [90].

2.2.6. The Monte-Carlo event generator PYTHIA

Monte Carlo event generators have various applications in experimental high-energy physics that reach from feasibility studies of analysis methods to ingredients for the signal extraction or background rejection. One event generator that is widely used for the simulation of lepton-lepton, lepton-hadron and hadron-hadron collisions is PYTHIA [56]. For PYTHIA, the event generation is based on LO matrix elements complemented by parton shower algorithms which sum the leading logarithmic contributions for all orders of the coupling constant. To achieve NLO accuracy, PYTHIA is often used in combination with the NLO matrix-element generator POWHEG [39, 95, 128]. The various parameters that are fed to PYTHIA simulations have been tuned to describe the data from the Large Electron-Positron collider (LEP) and the LHC. The standard tune since PYTHIA 8.2 is the Monash 2013 tune [148].
PYTHIA simulations for events with initial hard $2 \to 2$ QCD processes are used for several analysis steps discussed in this thesis, which is why the corresponding physical concepts shall be discussed in the following.

Generation of the hardest process   The event generation starts with the sampling of the hardest $2 \to 2$ process which can be calculated perturbatively as described in Sect. 2.2.2. To simplify further discussion, the particles that are entering the hard process shall be named *incoming* and those that are leaving the hard process shall be called *outgoing* hard-process particles. Each of the outgoing hard-process particles carries a momentum $\hat{p}_T$ that can be specified by the user and that sets the scale for the further evolution of FSR and ISR. Since the respective massless matrix elements diverge for $p_T \to 0$, a sensible minimal value of $\hat{p}_T$ needs to be chosen for the simulation of hard QCD processes [56].

Parton-shower evolution   After it has been decided on the type of hard process, ISR is evolved backwards in time starting from the incoming hard-process particles right back to the initial-state partons before any radiation activity. FSR is processed forwards in time reaching from the outgoing hard-process particles towards the partonic final state before hadronisation takes over. Apart from the direction of the temporal evolution, the mathematical description of FSR and ISR based on successive particle splittings is very similar. Splittings of partons a into partons b and c with a momentum fraction z and $z - 1$ are ordered with respect to the evolution variable $p_{\perp\text{ev}}^2$ which is defined via

$$p_{\perp\text{ev}}^2 = z(1-z)Q^2 \quad \text{for FSR and} \tag{2.2.12}$$

$$p_{\perp\text{ev}}^2 = (1-z)Q^2 \quad \text{for ISR.} \tag{2.2.13}$$

In that sense, splittings with a large $p_{\perp\text{ev}}^2$ are performed prior to splittings with a smaller $p_{\perp\text{ev}}^2$ and the maximum momentum $p_{\perp\text{ev,max}}^2$ is given by the momentum of the hard-process particles or by the transverse momentum of a previous branching. The evolution of ISR, FSR and MPIs

is fully interleaved in this ordering scheme: the process with the largest $p_{\perp\mathrm{ev}}^2$ is always generated first [56, 78, 143].

In the case of FSR, the probability $\Delta_{a\to bc}$ for a parton a to split into partons b and c at $p_{\perp\mathrm{ev}}^2 = p_{\perp\mathrm{spl}}^2$ is given by

$$\Delta_{a\to bc} = \Pi_a(p_{\perp\mathrm{ev,max}}^2, p_{\perp\mathrm{spl}}^2)\, \mathrm{d}\mathcal{P}_a(p_{\perp\mathrm{spl}}). \tag{2.2.14}$$

Thereby, $\mathrm{d}\mathcal{P}_a(p_{\perp\mathrm{spl}})$ is the probability that particle a splits with the intrinsic assumption that it still exists. $\mathrm{d}\mathcal{P}_a(p_{\perp\mathrm{spl}})$ is defined by the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) evolution [43, 89, 100] as the sum over all possible splitting products and by integrating over all possible scenarios of how the momentum is divided between them

$$\mathrm{d}\mathcal{P}_a(p_{\perp\mathrm{spl}}) = \frac{\mathrm{d}p_{\perp\mathrm{spl}}^2}{p_{\perp\mathrm{spl}}^2} \frac{\alpha_s(p_{\perp\mathrm{spl}}^2)}{2\pi} \sum_{b,c} \int_{z_{\min}(p_{\perp\mathrm{spl}}^2)}^{z_{\max}(p_{\perp\mathrm{spl}}^2)} P_{a\to bc}(z)\, \mathrm{d}z. \tag{2.2.15}$$

The corresponding splitting kernels $P_{a\to bc}(z)$ for $q \to qg$, $g \to gg$ and $g \to q\bar{q}$ branchings can be found in [56]. The no-emission probability $\Pi_a(p_{\perp\mathrm{ev,max}}^2, p_{\perp\mathrm{spl}}^2)$ that enters Eqn. 2.2.14 is called the *Sudakov factor* and measures the probability that particle a does not split for $p_{\perp\mathrm{spl}}^2 < p_{\perp\mathrm{ev}}^2 < p_{\perp\mathrm{ev,max}}^2$ and thus survives until $p_{\perp\mathrm{ev}}^2 = p_{\perp\mathrm{spl}}^2$. It can be calculated as

$$\Pi_a(p_{\perp\mathrm{ev,max}}^2, p_{\perp\mathrm{spl}}^2) = \exp\left(- \int_{p_{\perp\mathrm{spl}}^2}^{p_{\perp\mathrm{ev,max}}^2} \mathrm{d}\mathcal{P}_a(p_{\perp\mathrm{spl}})^2 \right). \tag{2.2.16}$$

The probability for an ISR branching is defined following the same principle as for an FSR branching. However, the abundance of parton a in the initial-state proton needs to be taken into account by considering the dependence on the corresponding PDFs in Eqn. 2.2.15 [56, 78, 143].

An important concept for the accurate description of gluon emissions by event generators is the concept of *colour coherence*: soft wide-angle radiation off the partners of a collinear parton pair can interfere coherently and can thus be described as being emitted by the pair as a whole [64, 133].

In PYTHIA, colour coherence is taken into account by the careful choice of $p_{\perp\mathrm{ev}}^2$ as the evolution variable and by considering *colour dipoles*. Exploiting the limit of infinite colours, a colour-flow scenario is assigned to each PYTHIA event history. The emission of one particle – the emitter – is always interlinked with its colour-connected partner which takes up the recoil. Emitter and recoiler thus form a colour dipole and radiation off this dipole depends on the dipole's properties [56, 64, 143].

The PYTHIA description of heavy flavours Q exhibits several peculiarities compared to the description of light flavours with respect to the following parton-shower processes:

- **Q→Qg radiation:** The dead-cone effect for heavy-flavour radiation necessitates *matrix-element corrections* to the parton-shower probability for gluon emissions such that the latter resembles the results from corresponding matrix-element calculations. In this respect, gluon emissions are weighted by the ratio of the matrix-element probability and the parton-shower probability which leads to an exact rescaling for the hardest gluon emission. A slightly modified but similar procedure is applied for further emissions. This procedure, however, does not modify the $g \to gg$ branchings of gluons that recoil from heavy flavours. A further mass-dependent suppression of $g \to gg$ dipole radiation is therefore implemented [56, 130].

- **g → QQ̄ branchings:** The default option for the description of heavy-flavour production via gluon splittings provokes a reweighting of the corresponding splitting kernel for massless quarks to resemble the branching ratio from ME calculations [109, 144]. The inclusive bb̄ production that is obtained in this way using the Monash 2013 tune is larger by about a factor $\frac{3}{2}$ than corresponding measurements for e^+e^- collisions [148]. These deviations may result from the choice of $p^2_{\perp \mathrm{ev}}$ as the renormalisation scale, the description of colour coherence effects or the value that is used for the bare quark mass. Corresponding studies show that a modified description of colour coherence effects has a significant effect on the bb̄ production [109, 130, 148].

Since the publication of PYTHIA version 8.3, event simulations can be performed using different shower algorithms. In particular, the Vincia shower algorithm [94] can be applied instead of the "simple shower" detailed above, which is the standard shower algorithm for PYTHIA 8 [56]. The Vincia algorithm describes the parton-shower history as a sequence of 2→3 particle branchings. This makes it an *antenna-based* parton shower with the individual colour-connected splitting partners forming the respective antennas that radiate as an entity. While the default dipole-based PYTHIA shower only approximately takes into account colour-coherence effects, they are explicitly considered by antenna-based showers [94]. As it was stated above, the description of colour coherence effects is interesting in terms of the $g \to$ bb̄ production rates. Interestingly, it was found in [107] that predictions for the beauty production at LHC energies and forward rapidities differ significantly for the Vincia shower in comparison to the PYTHIA simple shower.

Hadronisation   The process of hadronisation is described by the Lund String model. Within this model, the quark and the antiquark of a mesons are assumed to be connected by a tube of interacting gluons which is called *string*. Neglecting the Coulomb interactions of both quarks, the potential energy that is stored in the string grows linearly with the string lengths. The string constant, which is the respective proportionality constant, is given by $\kappa \approx 1\,\mathrm{GeV/fm}$. Considering massless quarks, new hadrons are produced by string breaks at a single vertex. On the other hand, massive quarks need to tunnel up to a certain distance at which the energy of the string balances their mass and momentum. The differential probability $\mathrm{d}\mathcal{P}$ for the tunneling process depends on the quark mass m and transverse momentum according to

$$\frac{1}{\kappa} \frac{\mathrm{d}\mathcal{P}}{\mathrm{d}^2 p_\mathrm{T}} \propto \exp(-\pi m^2/\kappa) \, \exp(-\pi p_\mathrm{T}^2/\kappa). \qquad (2.2.17)$$

This is why heavy quarks are only rarely generated in string breaking processes. In particular, the production of beauty quarks is suppressed with respect to the production of light flavours by a factor of 10^{-11} [56, 64].

2.3. Heavy-ion collisions and the QGP

In February 2000, it has been proclaimed at CERN that a new state of matter had been observed in the evolution of matter after a Pb–Pb collision: the quark-gluon plasma (QGP) [104, 129]. The findings of seven experiments at the SPS collider were combined to report "compelling evidence" [129] for the existence of a medium in which quarks and gluons are deconfined and can thus move freely. Since this statement was made, the heavy-ion physics community has put effort into unraveling the properties of the QGP and the corresponding phase transition from nuclear matter under normal conditions to the QGP. A short review on the present knowledge in this respect shall be given in this section. Following the focus of this thesis, special emphasis will be layed on so-called *hard probes*, meaning objects with large momentum, which can provide knowledge on the QGP properties.
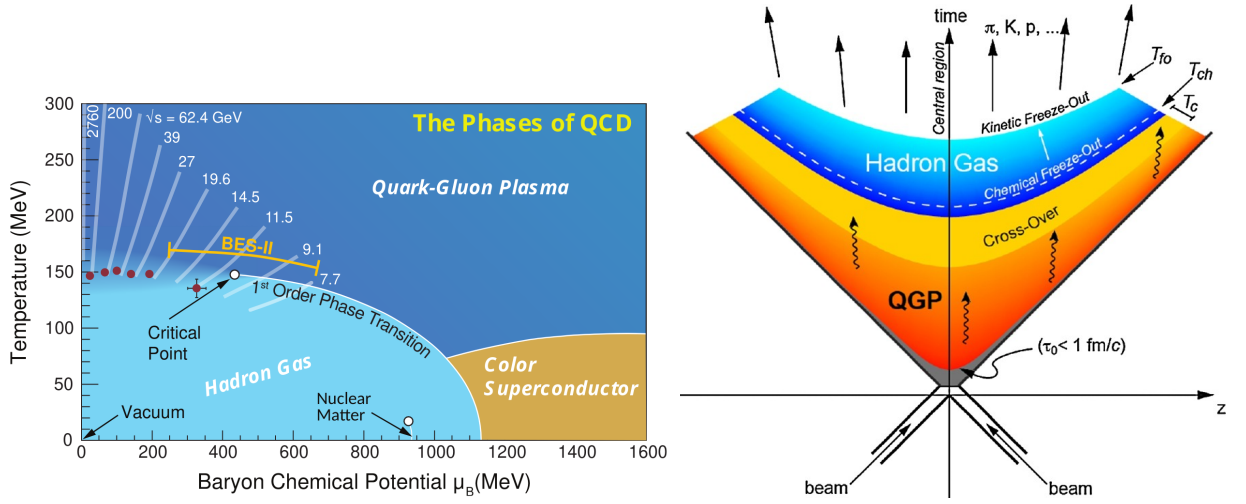
Figure 2.9. – **Left:** The phase diagram of QCD in dependence of the temperature T and the baryon chemical potential μ_{B} [66]. By the example of the beam-energy scan program (BES-II) at RHIC, it is shown how colliders with different centre-of-mass energies can investigate different regions of the diagram. Nowadays, one further energy needs to be added towards the outermost left at small μ_{B} as the LHC can probe the phase diagram with a center-of-mass energy of 5.02 TeV for Pb–Pb collisions. At the same time, FAIR will contribute in the future with high-precision measurements below 50 GeV [71]. **Right:** Schematic drawing of the chronology of different phases that matter undergoes after a high-energy heavy-ion collision [60].

Emergence and evolution of the QGP A QGP can be generated from normal nuclear matter in two ways: either through heating the latter to large temperatures or by compressing it to reach large baryon densities. Both of these transitions can be investigated using heavy-ion collisions as, depending on the centre-of-mass energy of the collision, a large variety of temperatures and densities can be accomplished. This gives the opportunity to investigate different regions of the phase diagram of QCD, which is shown in Fig. 2.9 in dependence of the temperature and the baryon chemical potential μ_{B} of the medium. In heavy-ion collisions at LHC energies ($\sqrt{s_{\mathrm{NN}}}=$2.76, 5.02 TeV) a QGP with small baryon densities (small μ_{B}) and large temperature can be studied. Experiments at other collider facilities, for example at RHIC ($\sqrt{s_{\mathrm{NN}}} = 7.7, \ldots, 200$ GeV) [66] or in future also at FAIR ($\sqrt{s_{\mathrm{NN}}} < 50$ GeV) [71], have the possibility to investigate regions of larger μ_{B} and smaller T [66, 158].

The nature of the phase transition from normal nuclear matter, which can be described as a hadron gas, to a state of deconfined quarks and gluons is still subject to investigations. Calculations from lattice QCD predict a smooth cross-over phase transition at low μ_{B}. At larger μ_{B}, a critical point is expected at which this cross-over transition changes to a first order phase transition. The search for this critical point and the examination of QGP properties close to it has motivated several experimental initiatives, of which the beam-energy scan program at RHIC [26] and the development of the Compressed Baryonic Matter experiment (CBM) [71] at FAIR are two examples [66].

Fig. 2.9 shows the evolution of matter after a high-energy AA collision, schematically. Following a short pre-equilibrium phase, the matter reaches a state of local thermal equilibrium – the aforementioned QGP [60]. The system expands and cools down until it undergoes a cross-over phase transition to a hadron gas. At the chemical freeze-out temperature T_{ch}, all particle yields are frozen and at the kinetical freeze-out with a temperature T_{fo} also elastic scatterings of the particles cease [60, 158].

Probing the properties of the QGP Investigations have shown that the QGP which is generated in high-energy collisions can be described by relativistic hydrodynamics. In particular, the
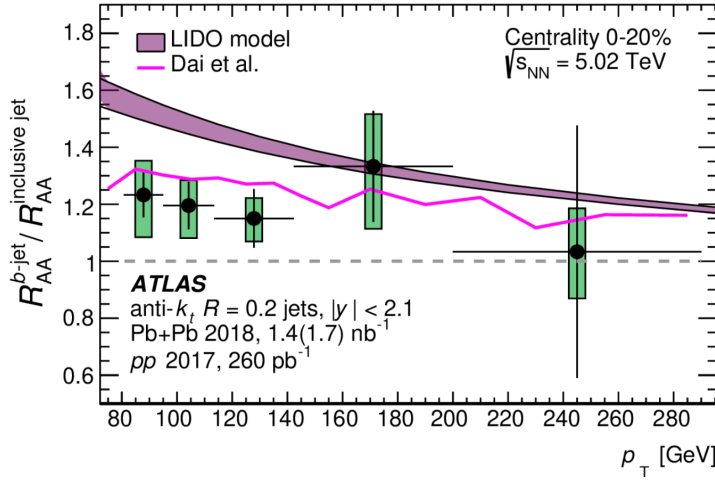
Figure 2.10. – The ratio of the R_{AA} for beauty jets over the R_{AA} for inclusive jets which has been measured by ATLAS in central Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02\,\mathrm{TeV}$ [123].

medium behaves like a perfect liquid in which the individual particles are strongly coupled. This strong coupling translates into collective behaviour of the particles within the plasma that can be measured using particle correlations. From these measurements as well as from particle transverse momentum spectra, it is possible to constrain the viscosity of the QGP, which is found to be close to the theoretical lower limit ($\eta/s = \frac{1}{4}\pi$) as it is the case for a perfect liquid [24, 65, 127].

Hard probes, like jets or heavy-flavour quarks, originate from hard parton scatterings that take place early in the history of matter evolution before the emergence of the QGP. Thus, they experience the evolution of the QGP – which is dominated by light quarks and gluons – and interact with the plasma constituents. The characteristics of these interactions are assumed to be imprinted on heavy-flavour observables as heavy quarks are not expected to fully thermalise in the medium. This makes them a perfect probe to study the QGP transport properties [54, 158]. When hard objects traverse the QGP, they interact with quarks and gluons in the plasma and energy is redistributed among all participants. They may lose energy via collisions and gluon radiation but they can also pick up momentum from the plasma. The energy loss via gluon radiation – which is the dominant energy loss mechanism at large momenta – proceeds in a similar manner as in the QCD vacuum described in Sect. 2.2.4. This means that the probability for gluon radiation depends on the partons' colour factor and mass. However, gluon radiation is modified in the presence of the QGP as the interaction with plasma constituents can trigger gluon radiation [57]. The examination of the difference between observables in pp and in AA collisions can thus provide insight into the QGP's properties.

One possibility to measure a corresponding deviation of particle properties in AA collisions with respect to pp collisions is the investigation of the *nuclear modification factor* which is defined as [158]

$$R_{\mathrm{AA}}(x) = \frac{\mathrm{d}N_{\mathrm{AA}}/\mathrm{d}x}{\langle N_{\mathrm{coll}}\rangle\, \mathrm{d}N_{\mathrm{pp}}/\mathrm{d}x}. \tag{2.3.1}$$

Thereby, $\mathrm{d}N_{\mathrm{AA}}/\mathrm{d}x$ is the jet or particle multiplicity distribution in dependence of the observable of interest x in AA collisions and $\mathrm{d}N_{\mathrm{pp}}/\mathrm{d}x$ is the corresponding distribution in pp collisions. The R_{AA} is scaled with respect to the number of binary nucleus-nucleus collisions $\langle N_{\mathrm{coll}}\rangle$ for the AA collisions that is calculated using the Glauber model [99]. In the past, the R_{AA} has usually been examined as a function of the transverse momentum of the object under investigation (i.e. $x \equiv p_{\mathrm{T}}$) [158].

The R_{AA} for various particle species has been investigated up to now and for many of these an R_{AA} smaller than one, meaning a suppression of particle production in AA with respect to

pp collisions, has been found at high p_T – a phenomenon which is called *jet quenching*. The amount of suppression differs for the individual particle species for reasons like a dependence of the energy loss of the involved quarks on mass and colour charge [25] or a dependence on hadron binding energies (e.g. sequential melting of quarkonia [142]).

Nowadays, significant effort is put into constraining the properties of energy redistribution via studying jet substructure observables and jets of different categories. In this context, substructure measurements show that jets in Pb–Pb collisions are broader than in pp collisions with the structure of the jet core being similar to the core in pp collisions. In particular, findings are consistent with the broadening of b jets being stronger than for inclusive jets [124]. In addition, b-jets have been found to be less suppressed in central Pb–Pb collisions than inclusive jets. This can be seen in Fig. 2.10 which provides the ratio of the b-jet R_{PbPb} over the R_{PbPb} of inclusive jets [1]. The difference can be explained by colour charge effects – inclusive jets are dominated by gluon jets – and potentially also the large mass of the beauty quark. A possibility to isolate mass-dependent effects in this context would be to investigate projections of the Lund Plane for b-jets with fully reconstructed b hadrons [125].

A thorough understanding of a QGP probe in the QCD vacuum needs to precede investigations on QGP properties in AA collisions. This further motivates the measurement of the b-jet production in pp collisions that is the major topic of this thesis as well as the simulation study of the Lund Plane for b jets in pp collisions which is described in Sect. 2.2.5.

3. Experimental facilities

For this thesis, data is analysed that has been taken by the ALICE experiment at the LHC in 2017. Therefore, the main details about both the LHC and ALICE will be described in this chapter.

3.1. The Large Hadron Collider

The LHC [92, 115, 117] is a circular particle accelerator with a circumference of 26.7 km that is located at CERN in Geneva. Since 2009 [154], high-energy pp, pA and AA collisions are evoked at four interaction points where the four major experiments ALICE, ATLAS, CMS and LHCb are located. All four experiments represent general-purpose detectors and study a wide range of observables. However, each experiment has a specific focus: The ALICE experiment is a dedicated heavy-ion experiment that has been designed to study the QGP with regards to the physics of the strong interaction. ATLAS and CMS constitute high-luminosity experiments which investigate rare high-p_T objects with the objective to examine the properties of the Higgs boson and search for physics beyond the SM. The low-luminosity experiment LHCb aims at searching for new sources of CP violation in rare decays of beauty and charm hadrons.

The LHC has been operating successfully for two full data-taking periods – LHC Run 1 (2009-2013) and LHC Run 2 (2015-2018) – intermitted by a first long shutdown period (LS1). After major upgrades of the detectors and the LHC machinery in the second long shutdown (LS2), a third run period LHC Run 3 has started in July 2022 and will continue until 2025 [114, 115, 152]. The main focus of the collision programme has been set on analysing pp collisions – corresponding data has been taken for centre-of-mass energies reaching from $\sqrt{s} = 0.9$ TeV at the beginning of Run 1 up to 13 TeV reached in Run 2. This pp programme was motivated by the search for and study of the Higgs boson and, as of yet, the announcement of its observation in 2011 by the CMS [74] and the ATLAS [4] experiment was the major discovery of LHC detectors.
Also heavy ions like $^{208}_{82}$Pb can be accelerated within the LHC rings and in this sense data for PbPb (2.76 and 5.02 TeV) and pPb collisions (5.02 and 8.16 TeV) have been recorded with the objective to investigate the properties of the QGP and separate them from cold nuclear matter effects [114].

As it is schematically shown in Fig. 3.1, the LHC exhibits two separate rings in which particles are accelerated in opposite directions. There are eight arc sections that keep the particles on a circular path using superconducting dipole magnets with a magnetic field of 8.33 T. These arcs are linked by eight straight sections and at the centre of four of these section, beams are crossed for particle collisions in the main experiments. Thereby, the beams are focused by an ensemble of focusing and defocussing quadrupole magnets.
The LHC was originally designed for accelerating protons up to 14 TeV with a luminosity of 10^{34} cm^{-2}s^{-1}; the so far maximum energy of 13.6 TeV was reached in July 2022 [152]. Before the particles gain their maximum energy in the LHC, they run through several pre-acceleration stages that are sketched in Fig. 3.2. For Run 2, protons have first been accelerated up to 50 MeV by the linear accelerator Linac2 before they have been injected into the Proton-Synchrotron Booster (max. 1.4 GeV), the Proton Synchrotron (PS, max. 25 GeV), the Super-Proton Synchrotron (SPS,
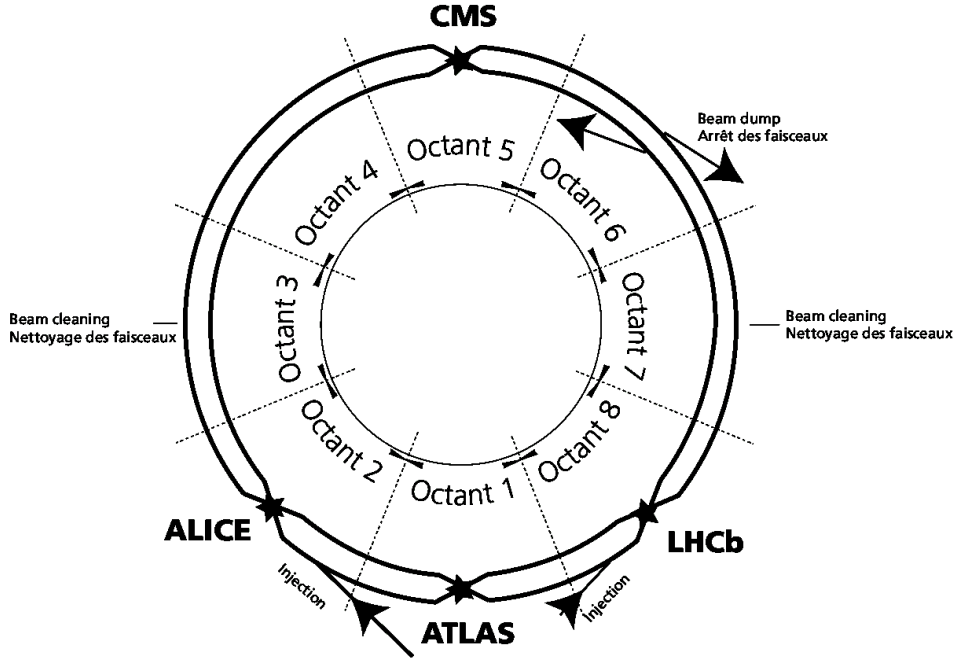
Figure 3.1. – Schematic picture of the circulation of the beams in the LHC rings [117, modified]. Both rings are build from eight arcs and eight straight sections. The ring segment that reaches from the center of an arc to the next arc center is termed an octant. Particle collisions are evoked by beam crossings at the center of the straight sections of octants one (ATLAS), two (ALICE), five (CMS) and eight (LHCb). The distance between the beam lines is not to scale.

max. 450 GeV) and finally the LHC. Run 3 comes with changes in the injection energies and a replacement of the linear accelerator Linac2 by its successor Linac4. For more information on the LHC machinery in Run 3 see e.g. [84].

Heavy-ions that are accelerated take a slightly different path: they first pass the linear accelerator Linac3 before they are injected into the Low Energy Ion Ring (LEIR) and later on in the PS. From there on, they undergo the same pre-acceleration steps as protons but with different intermediate energies due to their different charge-to-mass ratio.

3.2. The ALICE experiment

The ALICE detector system [8, 13] was designed with the special focus on uncovering the properties of the QGP that is generated in heavy-ion collisions. This objective has determined the choice and characteristics of the 18 subdetectors of ALICE which are well suited to operate in an environment with high particle densities of several thousand particles per unit rapidity. In this sense, high-granularity detectors have been preferred above detectors that can take data at high read-out rates. ALICE can reconstruct particle tracks with a p_T resolution better than $\sigma_{p_T}/p_T = 2\,\%$ for track momenta between $150\,\text{MeV}/c$ and $20\,\text{GeV}/c$ [13, 17]. With this precision down to low p_T, ALICE is well suited to complement high-p_T jet measurements by ATLAS and CMS. Although ALICE is specialised for the examination of PbPb collisions, it also takes data in pp and pA collisions as a reference for heavy-ion collisions whereby most analyses concentrate on taking data at midrapidity in the region of the largest energy density.

A schematic picture of the ALICE experiment and its subdetectors for LHC Run 2 can be found in Fig. 3.3. The whole detector system comprises a volume of $16 \times 16 \times 26\,\text{m}^3$ and can be divided into the central barrel detectors and the forward muon spectrometer. The central barrel
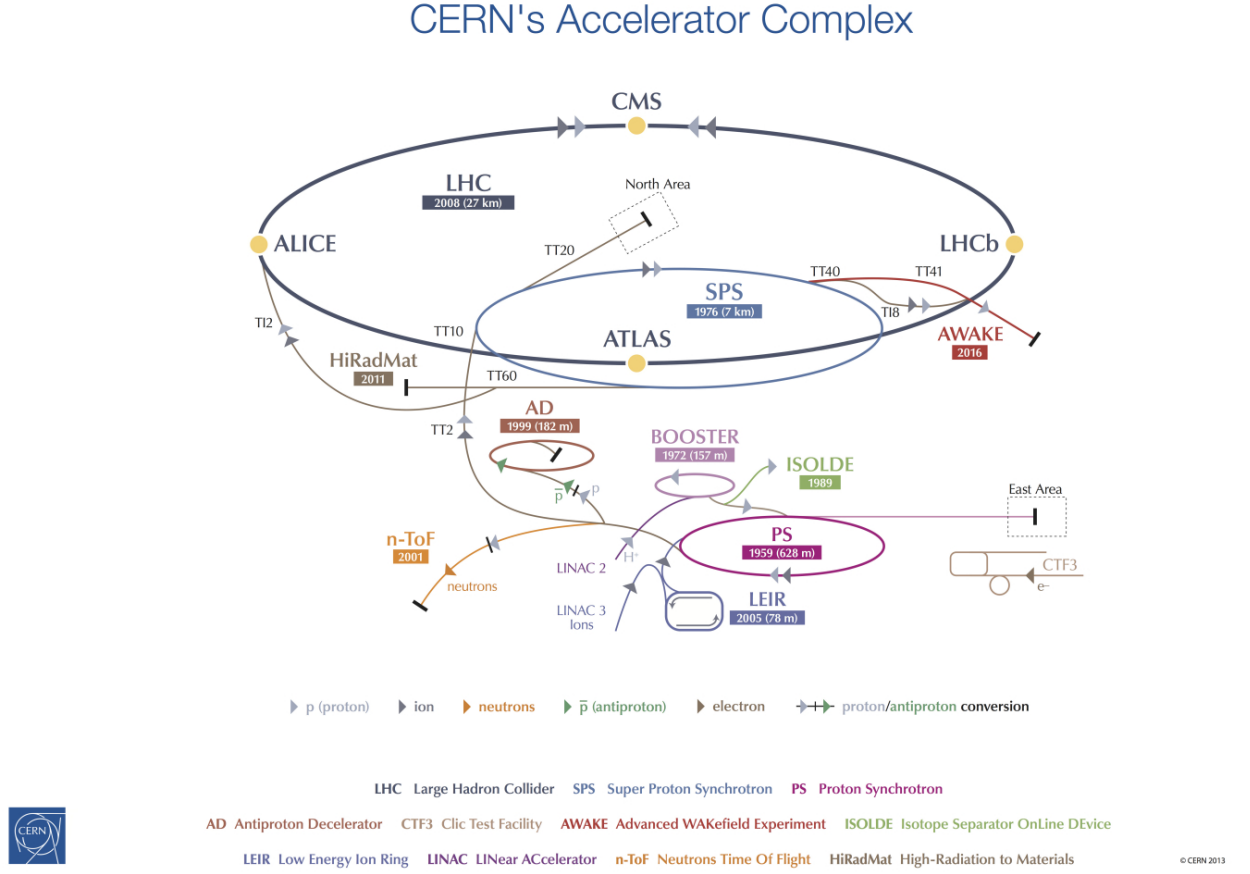
Figure 3.2. – The CERN accelerator complex for LHC Run 2 [101]. For LHC Run 3, the linear accelerator Linac2 that is used as a first acceleration step for protons is replaced by Linac4, see [126].
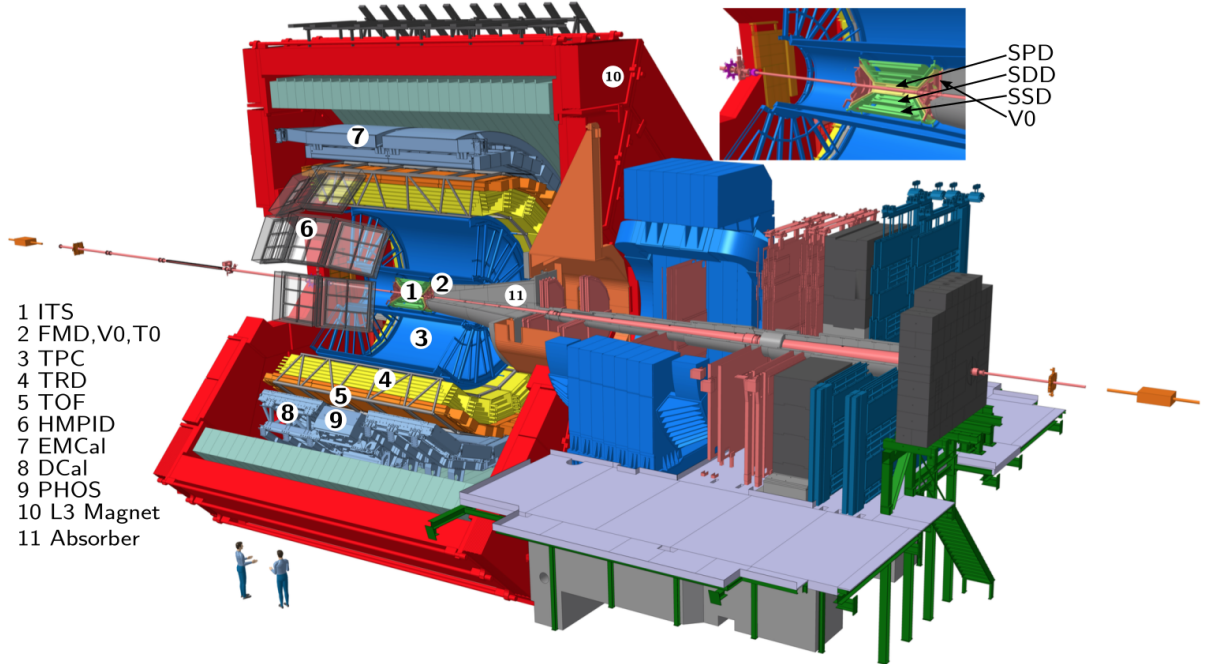

Figure 3.3. – The ALICE Experiment with the subdetector system for LHC Run 2. The locations of the central barrel detectors as well as the solenoid magnet are indicated [150, modified].

detectors are embedded in a solenoid magnet with a magnetic field of 0.5 T which permits the determination of the particles' charge and momentum via their curvature in the magnetic field – important ingredients for the particle identification (PID).

Fig. 3.4a shows the ALICE coordinate system as it has been defined in [55] and as it is used for the construction of measurement observables. The coordinate origin is given by the nominal beam interaction point. The z-axis points along the beam pipe in the opposite direction of the muon spectrometer such that the xy-plane is perpendicular to the beam direction. The azimuthal angle φ describes the angular distance of a particle track towards the x-axis in the xy-plane whereas the polar angle θ measures the angular distance with respect to the z-axis in the zy-plane.

3.2.1. The ALICE subdetectors

During the LS2, several ALICE detectors were subject to major upgrades. However, since data from Run 2 is analysed in this thesis, the status of the subdetectors for this time period will be described in the following. A short overview on the most important optimisations for Run 3 with respect to the heavy-flavour measurements will be given at the end of this subsection.

The subdetectors that are most important with respect to beauty-jet tagging with ALICE are the Inner Tracking System (ITS), the Time Projection Chamber (TPC) and the V0 detector. This is why, the latter detectors will be described in more detail while all other detectors are mentioned only briefly.

The Inner Tracking System (ITS) is especially important for the heavy-flavour identification as it reconstructs not only the primary vertex of the collision but also secondary decay vertices of long-lived particles like beauty hadrons. For the secondary-vertex finding, an adequate resolution of the transverse impact parameter – the track's distance to the primary vertex in the xy-plane (see Sect. 4 for precise definition) – is essential. The ITS thereby reaches a resolution of better than $80\,\mu$m ($20\,\mu$m) for charged pions with $p_\mathrm{T} > 1\,\mathrm{GeV}/c$ ($p_\mathrm{T} > 10\,\mathrm{GeV}/c$) [11]. At the same time, it complements the TPC with respect to charged-particle tracking by reconstructing low-p_T particles that do not reach the TPC and by providing tracking information in the TPC's dead regions. In addition, the ITS is utilised for the PID of low-p_T ($p_\mathrm{T} < 100\,\mathrm{MeV}/c$) particles since these can not be identified by the TPC.

As it can be seen in Fig. 3.4b, the detector system consists of six separate layers which detect particles with increasing granularity the closer they are situated to the beam pipe. The two innermost layers are Silicon Pixel Detectors (SPD) that embrace the beam pipe at a radius of 4 cm. These are surrounded by two layers of Silicon Drift Detectors (SDD) and two layers of Silicon Strip Detectors (SSD) with the latter reaching up to a radius of 43 cm. Both, the SSD and the SDD exhibit analog read-out to enable PID using energy-loss measurements.

The ITS covers the full azimuthal angle and has a rapidity coverage of $|\eta| < 0.9$ for primary vertices with $z = \pm 6$ cm relative to the nominal interaction point [11, 85].

The Forward Multiplicity Detector (FMD) is a silicon strip detector which is used for the offline estimation of the charged particle multiplicity at rapidities ($-3.4 < \eta < -1.7$ and $1.7 < \eta < 5.0$ [80]).

The V0 detector consists of two scintillator counters – V0A and V0C – which are situated on either side of the interaction point and have a rapidity coverage of $2.8 < \eta < 5.1$ and $-3.7 < \eta < -1.7$, respectively. These provide a minimum bias trigger for pp and AA collisions. Usually, both V0A and V0C are asked to deliver a positive trigger decision ("V0AND" decision) for an event to be accepted in this respect. The V0 detector also provides a centrality trigger for AA collisions and the possibility to reject background from

(a) ALICE coordinate system.

(b) ALICE Inner Tracking System (ITS).

(c) ALICE Time Projection Chamber (TPC).
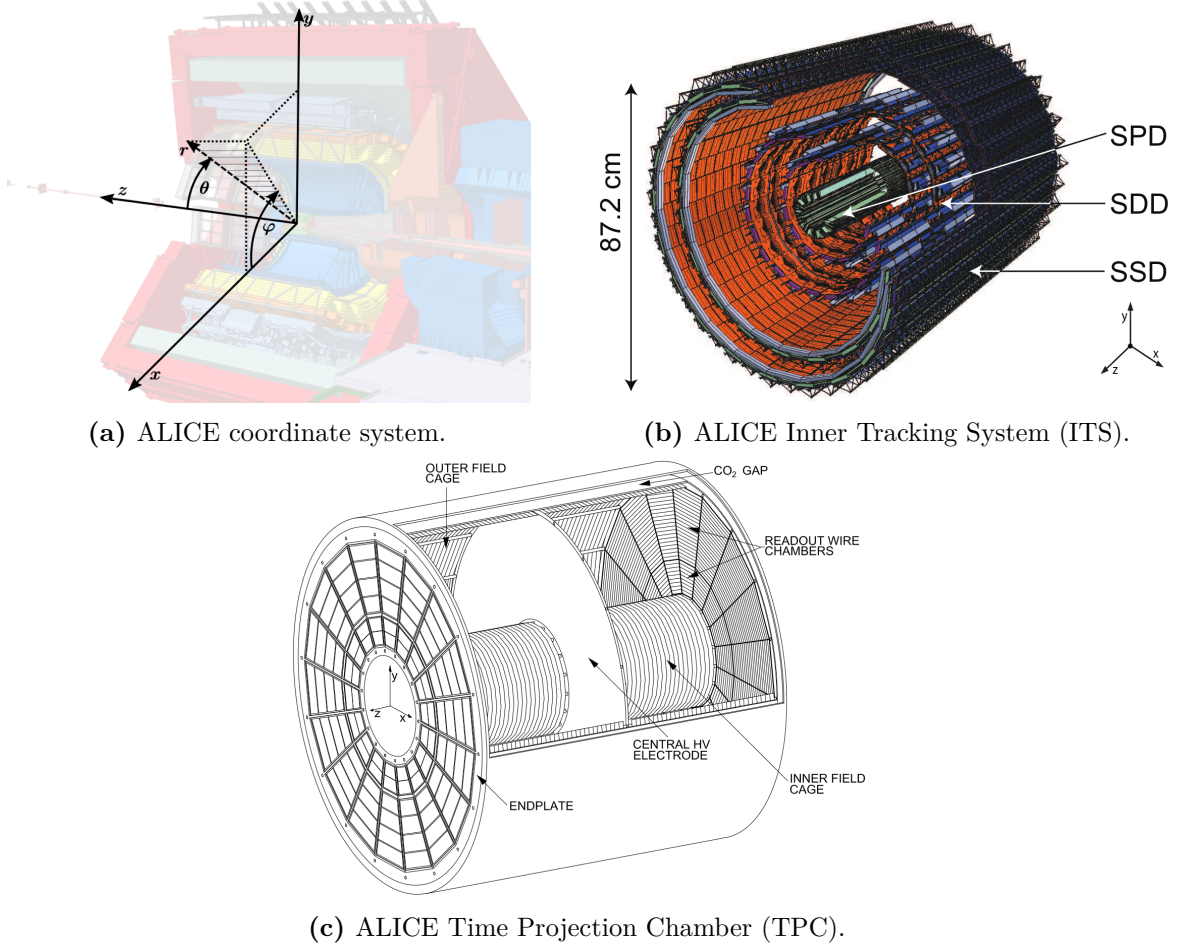
Figure 3.4. – **a)** The ALICE coordinate system [55], the azimuthal angle φ and the polar angle θ [150, modified]. See text for more information. **b)** The six layers of the ALICE Inner Tracking System. From the innermost to the outermost detectors these are the Silicon Pixel Detectors (SPD), the Silicon Drift Detectors (SDD) and the Silicon Strip Detectors (SSD) [9]. **c)** Scheme of the ALICE Time Projection Chamber [42].

proton-residual gas interactions or from background triggers in the muon spectrometer [8, 80].

The T0 detector exhibits two arrays of Cherenkov counters – T0A and T0C – on either side of the interaction point with a rapidity coverage of $4.6 < \eta < 4.9$ and $-3.3 < \eta < -3.0$. It provides a start signal for the Time of Flight (TOF) detector with a precision of $50 \, \text{ps}$ as well as a wake-up signal for the Transition Radiation Detector (TRD). At the same time, it serves as a minimum bias and multiplicity trigger and a trigger which rejects invalid vertex positions [8, 80].

The Time Projection Chamber (TPC) is the main tracking detector of ALICE and, together with the ITS, provides PID via energy-loss measurements with the most significant separation power for particles with $p_\text{T} < 500 \, \text{MeV}/c$. The energy-loss resolution that is achieved in pp collisions is about $5.2 \, \%$.

The detector, which is shown schematically in Fig. 3.4c, constitutes a hollow cylinder with a length of $500 \, \text{cm}$, an inner radius of $85 \, \text{cm}$ and an outer radius of $250 \, \text{cm}$. In the direction of the beam pipe, it surrounds the ITS whereas towards the outside it is enclosed by the TRD. The TPC contains about $90 \, \text{m}^3$ of counting gas which is a mixture of the noble gas Neon and CO_2 as a quencher. The cylinder volume is divided into two parts by a central electrode which generates an electric field with respect to the cylinder end plates. Charged particles that traverse the detector, ionise the detector gas and the corresponding ionisation electrons drift along the electric field towards the cylinder ends where they enter the read-out chambers. These are Multi-Wire Proportional Chambers (MWPC) with pad read-out that first amplify and then collect the charge signal via induction on the pad plane. The $xy-$coordinates of a particle's position are derived from the coordinates of the read-out pad that the produced electrons enter and the z-coordinate can be calculated from the electron's time of flight.

As the ITS, the TPC has an acceptance of 2π in azimuthal angle and a rapidity coverage of $|\eta| < 0.9$ [13, 17, 42].

The Transition Radiation Detector (TRD) with a rapidity coverage of $-0.84 < \eta < 0.84$ separates electrons and pions for momenta above $1 \, \text{GeV}/c$ via the detection of transition radiation. The latter is emitted by both particle species when they are passing boundaries of different permittivity in the radiator material in dependence of their Lorentz factor γ [8, 79].

The Time of Flight Detector (TOF) is an array of Multigap Resistive-Plate Chambers that provides PID for pions, kaons and protons at intermediate momenta with $|\eta| < 0.9$ [8, 38]. Since open heavy-flavour mesons have significant branching fractions into kaons, the information from the TOF can be used to identify D mesons based on the invariant mass of their decay products as it is done e.g. in [19].

The High-Momentum Particle Identification Detector (HMPID) is a Ring Imaging Cherenkov Counter with a pseudorapidity coverage of $|\eta| < 0.6$ that extends the PID capability of the other ALICE PID detectors for pions, kaons and protons to momenta larger than $1 \, \text{GeV}/c$ [8, 132].

The Electromagnetic Calorimeter (EMCal) is a Shashlik calorimeter built from layers of lead and scintillator material. It provides a jet trigger and enables ALICE to reconstruct not only charged but also inclusive jets as it detects electrons and high-p_T photons. The detection of photons also allows for the reconstruction of short-lived neutral mesons. The detector is situated between the TOF and the ALICE magnet coils and has a rapidity coverage of $|\eta| < 0.7$ and a coverage in azimuthal angle of $107°$ [13, 81].

The Dijet Calorimeter (DCal) is constructed using the same detector design as the EMCAL and is mounted on the opposite side of the beam pipe with respect to the EMCal. Thus, the DCal complements the EMCal making the investigation of hadron-jet and di-jet measurements possible. As the EMCal, it covers a rapidity region of $|\eta| < 0.7$ but has a lower azimuthal angle coverage of 60° [41].

The PHOton Spectrometer (PHOS) is an electromagnetic calorimeter built of lead-tungstate crystals with a rapidity coverage of $|\eta| < 1.2$ that is used to identify low-p_{T} direct photons as well as π^0 and η mesons up to a momentum of $10\,\mathrm{GeV}/c$ [122].

For LHC Run 3, the ITS has been replaced by ITS2 which achieves an impact parameter resolution that is improved by a factor of about three for charged particles with $p_{\mathrm{T}} = 1\,\mathrm{GeV}/c$ and a factor of about five for $p_{\mathrm{T}} > 10\,\mathrm{GeV}/c$. This progress has been made possible by decreasing the diameter of the beam pipe in the ALICE interaction region, by reducing the material budget and decreasing the pixel size of the ITS detector layers [11, 137]. To match the design goal of taking data at larger readout rates, the readout mode of most detectors, and in particular of the TPC, have been changed from triggered mode to continuous readout. In case of the TPC, this necessitated the replacement of the front-end and read-out electronics as well as a replacement of the former MWPCs for gas amplification by Gas Electron Multipliers (GEMs) [27]. Both, the improvement of the impact-parameter resolution as well as the increase in statistics due to continuous readout will allow ALICE to fully reconstruct b hadrons in data from LHC Run 3. This full reconstruction was not feasible for data from LHC Run 2 [98]. However, the beauty production can also be studied in different ways that are discussed in Sect. 4. In particular, jets initiated by beauty quarks can be identified by exploiting the long lifetime of the weak decays of beauty hadrons, which is the main topic of this thesis.

3.2.2. Track reconstruction with ALICE

When charged particles traverse the ALICE detector, they induce electric signals in different subdetectors. The information from these *hits* can be used to reconstruct the particles properties like charge and momentum as well as their trajectories – the *tracks*. Track finding [13, 118] in ALICE starts with a first estimation of the primary interaction point based on the vertex with the largest number of SPD *tracklets*. Thereby, SPD tracklets correspond to matching signals in both SPD layers. The further reconstruction procedure is shown in an illustrative manner in Fig. 3.5 and can be divided into a first inwards propagation ("1st path" in Fig. 3.5), an outwards propagation ("2nd path") and a final inwards propagation ("3rd path"). The tracks are propagated through the detector volume using the Kalman filter technique [97]. For this iterative approach, first a prediction for a subsequent cluster is calculated based on the information of previous iterations and if a suitable cluster is found in the predicted area, all track parameters and the track's covariance matrix are updated for the newly gained constraints.

There are two ALICE reconstruction modes: the reconstruction of *TPC-ITS global tracks* and *ITS standalone tracking*. Whereby the first mode performs seed finding in the TPC and passes through all steps as mentioned above, the second method omits the first inwards propagation from the TPC towards the ITS and performs seed finding in the first two ITS layers. In this way, information about low-p_{T} tracks that do not reach the TPC as well as tracks that pass dead areas of the TPC are not lost but complemented by the ITS standalone tracking.
For ITS-TPC global tracks, the first reconstruction path starts at the outer TPC radius where seeds for primary tracks are constructed from two clusters and the estimated primary vertex. Seeds for secondary tracks are obtained from three clusters and all seeds are propagated towards the inner TPC radius. They then serve as seeds for track finding in the ITS whereby every TPC track can induce several track hypotheses in the ITS. After every ITS track candidate has been propagated towards the inner radius of the ITS, the TPC-ITS track hypothesis with the best $\mathcal{X}^2$

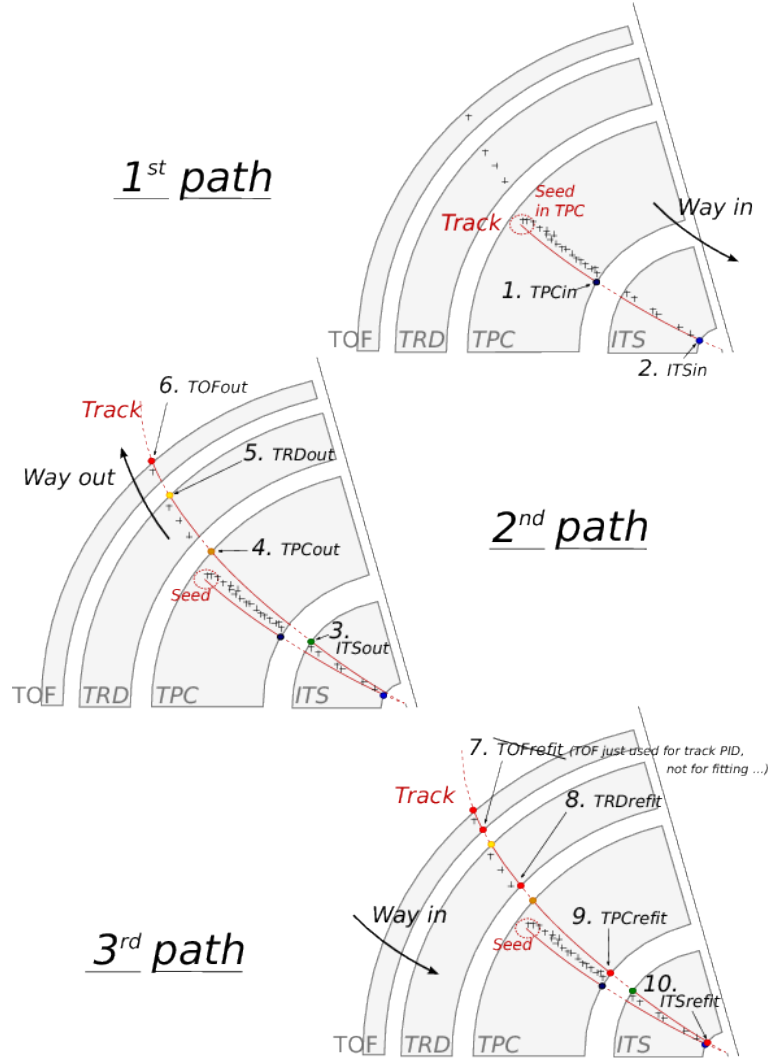**Figure 3.5.** – Schematic diagram of the track reconstruction process in ALICE which consists of a first inwards propagation ("1st path"), an outwards propagation ("2nd path") and a final inwards propagation ("3rd path") of cluster information. The numbers from 1 to 10 correspond to bits that are set for every successful completion of a substep. See text for more information. Figure taken from [118, 120]
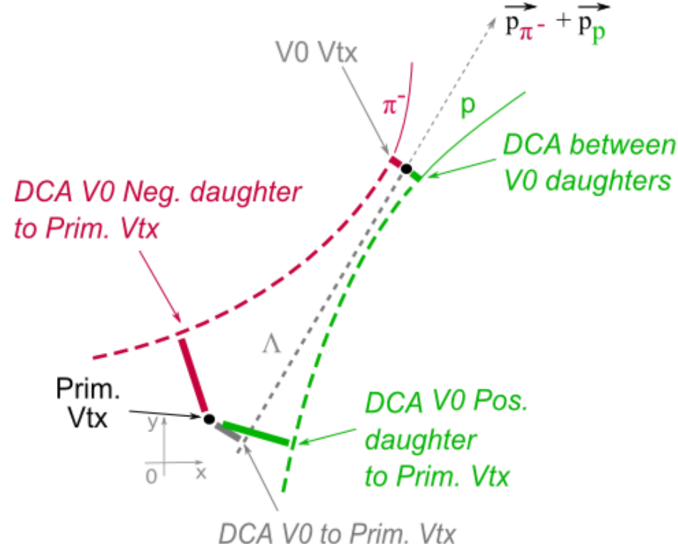
Figure 3.6. – Schematic representation of the decay $\Lambda \rightarrow p\pi^-$ as an example for the decay of a V^0 candidate [118, 119]. The secondary vertex of a V^0-hadron decay is well separated from the primary vertex which is why the corresponding decay tracks exhibit a large distance of closest approach (DCA) to the primary vertex.

value is selected.

In case of ITS standalone tracking, seeds are calculated from the primary vertex and two SPD clusters. These seeds as well as ITS-TPC global tracks are then propagated towards the ITS outer radius and matched to TRD and TOF tracklets, if possible. The propagation of the tracks to EMCal, PHOS and HMPID are not used for the estimation of the track parameters but are stored for further analysis.

As a last step, tracks are propagated back inwards starting from the outer TPC radius and all track parameters are updated for this final refit of previously found clusters.

As soon as the track reconstruction is concluded, the interaction vertex is updated from the information of the TPC-ITS global tracks. It is searched for γ conversions in the detector material as well as secondary decay vertices. In particular, pairs of unlike-sign tracks with a distance of closest approach to the primary vertex of larger than 0.5 mm are paired to so-called V^0 *candidates*. Among the V^0 candidates, there are tracks from long-lived neutral particles such as K_s^0 or Λ^0 mesons of which the secondary vertex is well displaced from the primary vertex and that decay into charged daughter tracks as shown schematically in Fig. 3.6. Secondary vertices of heavy-flavour mesons are too complex to be calculated at this stage which is why their reconstruction is done as offline analysis.

4. Measuring beauty – a short review

The intentions for investigations of the beauty production in small collision systems reach from testing predictions by pQCD, understanding the fragmentation of heavy-flavour hadrons up to establishing a baseline for measurements in heavy-ion collisions. There are several observables and analysis strategies that provide information about the properties of beauty production. Apart from analysing fully reconstructed beauty hadrons, it is possible to look at beauty-decay electrons (e.g. see [12, 131]), non-prompt charm hadrons (e.g. see [19]) or beauty jets (e.g. see [21]). Being without the possibility to reconstruct beauty hadrons in LHC Run 1 and Run 2, ALICE has up to now focused on the latter three possibilities.

The majority of available beauty analyses exploits a characteristic of beauty-hadron decays that arises from their long lifetime: As shown in Fig. 4.1, the *secondary vertex* at which b hadrons decay is well displaced from the primary vertex of the collision. One example for the definition of a suitable observable based on this geometrical signature is the *track transverse impact parameter* d_0. The latter is the projection of the distance of closest approach of a track towards the primary vertex onto the plane transverse to the beam axis; for ALICE the xy-plane as it can be seen in Fig. 4.1. This observable can further be divided by the resolution σ_d of d_0 and signed according to the scalar product of the impact-parameter vector $\vec{d_0}$ and the vector $\vec{e}_{\text{Jet}}$ of the jet axis. The final observable that is obtained like this has already been applied for beauty measurements at the Large Electron Positron (LEP) collider [10] and is called the *signed transverse impact parameter significance*

$$Sd_{xy} = d_0 \, \text{sign}(\vec{d_0} \cdot \vec{e}_{\text{Jet}})/\sigma_d. \tag{4.0.1}$$

Thereby, especially the consideration of a signed observable leads to a significant increase of separation power. Tracks originating from the secondary vertex of beauty decays tend to be aligned with the direction of the jet axis and thus, they are assigned a positive sign with a larger probability than tracks that originate from the primary vertex. Fig. 4.1 exemplarily shows the Sd_{xy} distributions for jets of different flavours in the momentum range of $30 < p_{\text{T,ch jet}} < 40 \, \text{GeV}/c$. The shape of the left-hand side ($Sd_{xy} < 0$) of the distributions is driven by detector effects that are dominated by the impact-parameter resolution. By construction, the right-hand side ($Sd_{xy} > 0$) carries information on the lifetime of the primary hadrons such that Sd_{xy} distributions for beauty jets exhibit larger tails towards positive values than light-flavour or charm jets.

Observables like the impact parameter can serve as input to beauty analyses in different ways. A first possibility are statistical methods that estimate the contribution of beauty objects to inclusive distributions via *template fits*. In this case, probability density functions (p.d.f) called templates are obtained for all quark flavours and their weighted sum is fit to the data. The relative contribution of the individual flavours is retrieved as fit result (e.g. see [93, 131]). For several applications, it is however favourable to identify beauty objects on a per-jet basis. In these cases, individual beauty objects are *tagged* using cuts on either single observables (e.g. see [21]) or decisions by complex machine-learning (ML) algorithms that process large samples of different input observables. The limited efficiencies and purities of these taggers need to be taken into account by respective corrections that in turn are often determined via template fits(e.g. see [5, 19, 141]).

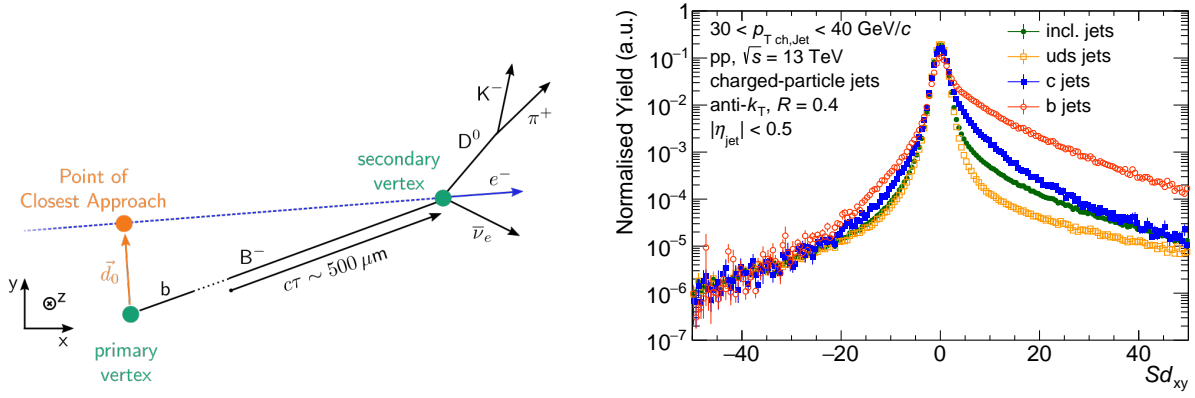Since the analysis presented in this thesis investigates beauty jets in pp collisions, corresponding

4. Measuring beauty – a short review

Figure 4.1. – **Left:** Sketch of the decay chain of a B^- meson to picture the characteristic geometry of the decays of long-lived b hadrons. The decay vertex, which is termed secondary vertex, is well displaced from the primary vertex of the collision. Beauty events can thus be identified by the appearance of tracks with large transverse impact parameter $d_0 = |\vec{d_0}|$. See text for more information. **Right:** Distributions of the signed transverse impact-parameter significance Sd_{xy} for jets of different jet flavour with $30 < p_{\mathrm{T,ch\ jet}} < 40\,\mathrm{GeV}/c$.

measurements by LHC experiments shall be reviewed shortly in the following. The summary starts from a measurement by ALICE in $5.02\,\mathrm{TeV}$ pp collisions [21], since the respective analysis strategy has served as a baseline for this analysis. The measurement addresses charged-particle beauty jets with radius $R = 0.4$ and a transverse momentum of $10 < p_{\mathrm{T,ch\ jet}} < 100\,\mathrm{GeV}/c$, which illustrates the unique reach of ALICE down to low values of $p_{\mathrm{T,ch\ jet}}$. The final b-jet spectrum is a combination of results from two different tagging and correction procedures referred to as *IP method* and *SV method*. For the IP method, jets that contain tracks with a large Sd_{xy} are selected and the performance is estimated via template fits to jet probability distributions (see Sect. 5.4.1 for definition of the jet probability) in data. The method is operated at a tagging efficiency of about $50\,\%$ to $70\,\%$ and a purity of about $40\,\%$ to $30\,\%$ from low to large $p_{\mathrm{T,ch\ jet}}$ (see Fig. 4.2). In case of the SV method, jets with a large two-dimensional decay length significance are tagged. This observable is defined as the ratio of the distance between the primary vertex and the secondary vertex in the plane transverse to the beam axis divided by the respective resolution. The efficiency for the SV method is directly estimated from PYTHIA 8 simulations while the purity is obtained from POWHEG simulations as well as template fits to distributions of the invariant mass of the secondary vertex. The method reaches a light-flavour misidentification probability of $0.1\,\%$ to $1\,\%$ from low to large $p_{\mathrm{T,ch\ jet}}$ for a tagging efficiency of $30\,\%$.

ALICE data has also been used to measure beauty jets in $7\,\mathrm{TeV}$ pp collisions [93]. This analysis pursues a statistical approach in a way that template fits to Sd_{xy} distributions are performed and the b-jet contribution for $5 < p_{\mathrm{T,ch\ jet}} < 100\,\mathrm{TeV}$ is directly obtained from these fits.

Apart from these ALICE measurements, ATLAS [3] and CMS [73] have obtained inclusive beauty-jet spectra for $7\,\mathrm{TeV}$ pp collisions with jet transverse momenta of $18 < p_{\mathrm{T,jet}} < 200\,\mathrm{GeV}/c$ in case of CMS and $20 < p_{\mathrm{T,jet}} < 400\,\mathrm{GeV}/c$ in case of ATLAS. Whereas these first b-jet measurements of LHC experiments were selecting jets based on a single observable – the three-dimensional decay-length significance – current offline tagging algorithms of both collaborations employ ML techniques with several input observables [5, 141]. In this sense, the b-jet cross section in $13\,\mathrm{TeV}$ pp collisions for a jet transverse momentum larger than $74\,\mathrm{GeV}/c$ up to several TeV has been obtained from CMS data by applying a multilayered neural network [77]. Various complementary methods have been developed by ATLAS and CMS for the determination of the tagging performance like the efficiency estimation based on muon-jet or $t\bar{t}$ dijet samples in data. Using these techniques, both experiments are able to identify beauty jets with a tagging efficiency of about $70\,\%$ for a light-flavour misidentification probability of about $1\,\%$. In addition to the offline tagging algorithms, ATLAS and CMS exhibit dedicated b-jet online triggers.
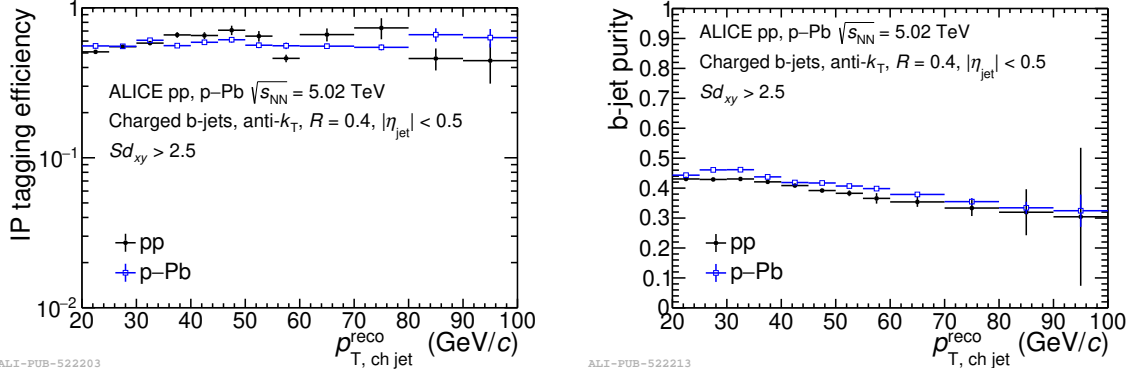
Figure 4.2. – The efficiency (left) and purity (right) for the tagging of beauty jets with the IP method of ALICE for 5.02 TeV pp collisions [21]. Note the logarithmic scale of the ordinate for the left figure.

The b-jet measurements discussed above have been compared to theoretical predictions. From these comparisons, it can be learned that NLO calculations using POWHEG in combination with PYTHIA 6 or 8 for the modeling of fragmentation describe the measured cross section and the ratio of beauty over jets of inclusive flavours within the uncertainties [3, 21, 77]. An example for the comparison of the b-jet cross section and the b-jet fraction in 5.02 TeV pp collisions to POWHEG+PYTHIA 8 simulations is provided in Fig. 4.3.

In contrast, *standalone* simulations with the LO MC event generator PYTHIA overestimate the beauty-jet production [3, 73, 77]. This can, for instance, be observed for the cross section of inclusive b jets at midrapidity measured by the CMS collaboration for pp collisions at $\sqrt{s} = 7$ TeV (see Fig. 4.4). At midrapidity and for transverse momenta of about $20\,\text{GeV}/c$, the prediction by PYTHIA 6 for the b-jet cross section is larger than the data by about a factor of two; the predictions are consistent with the data for momenta larger than $50\,\text{GeV}/c$. A similar effect can also be observed in Fig. 4.5 for the ratio of PYTHIA 8 simulations with the CMS underlying-event tune CUETP8M1 [113] over the measured b-jet cross section in 13 TeV pp collisions. In this case, the PYTHIA simulation is larger by a factor of 1.5 at about $80\,\text{GeV}/c$. On the other hand, the b-jet fraction of the latter measurement is well described by the simulation. Furthermore, measurements by the ALICE collaboration for charged-particle jets of *inclusive* flavours in pp collisions at $\sqrt{s} = 13$ TeV have been compared to POWHEG+PYTHIA 8 and PYTHIA 8 simulations. In the range of the transverse jet momentum that is interesting for this thesis, PYTHIA 8 overestimates the cross section at low $p_{\text{T,ch jet}}$ by about a factor of 1.5 while POWHEG+PYTHIA 8 simulations are compatible with the data [22].

In conclusion, previous measurements have show that the consideration of NLO processes is necessary to describe data for the cross section of b jets and jets of inclusive flavours in the transverse momentum range that is interesting for this thesis. In contrast, the relative abundance of b jets in comparison to jets of inclusive flavours is described by PYTHIA simulations.

The analysis that is presented in this thesis contributes to the collection of b-jet measurements by extending the CMS results for 13 TeV pp collisions down to $10\,\text{GeV}/c$. To do so, the analysis steps of the IP method by ALICE have been refined to improve in particular the precision of the performance estimation at low $p_{\text{T,ch jet}}$. As it was described in Sect. 2.2.4, the region of low momenta is of particular interest for investigations related to the radiation pattern of heavy flavours.
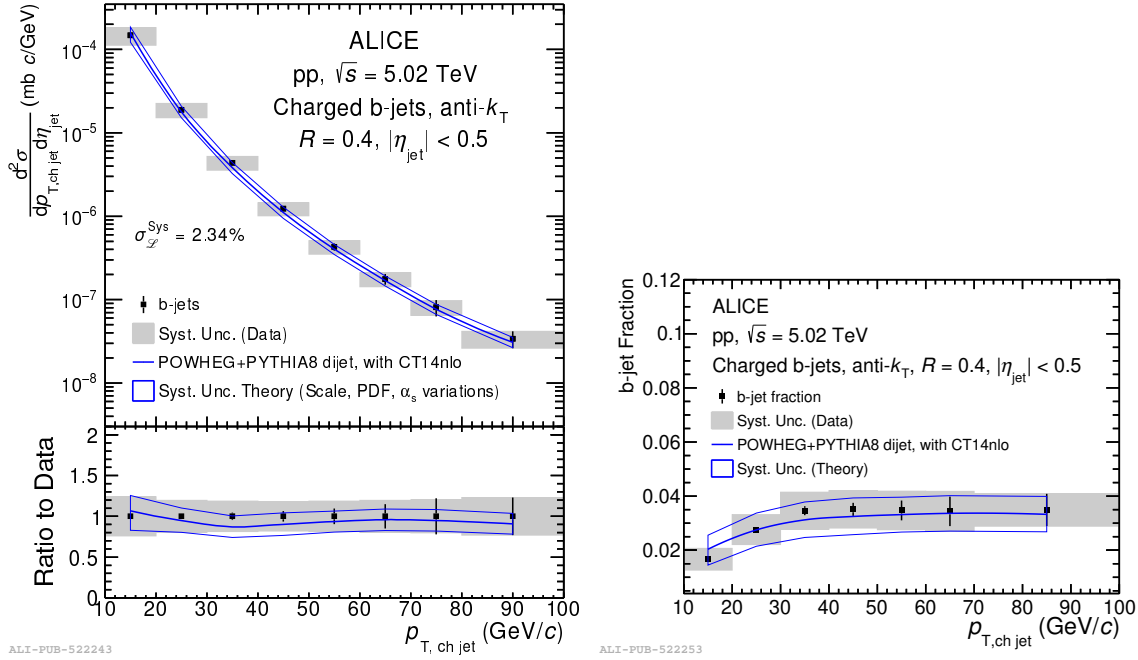
Figure 4.3. – The charged b-jet cross section for $5.02\,\mathrm{TeV}$ pp collisions that has been measured by ALICE is compared to predictions by POWHEG+PYTHIA 8 simulations. **Left:** The b-jet cross section [21]. **Right:** The ratio of the b-jet cross section over the cross section for jets of inclusive flavours at the same centre-of-mass energy [21].
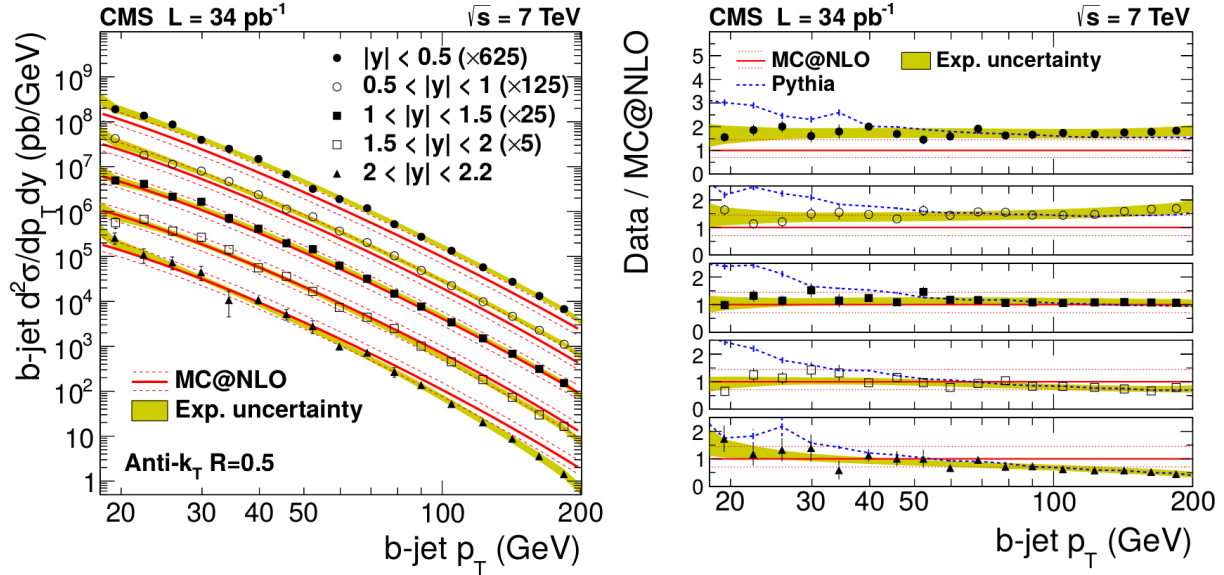
Figure 4.4. – The inclusive b-jet cross section in $7\,\mathrm{TeV}$ pp collisions that has been measured by CMS is compared to predictions by MCNLO [95, 96] and PYTHIA 6. **Left:** The b-jet cross section for different rapidity intervals [73]. **Right:** The ratio of the data over the theoretical predictions [73].
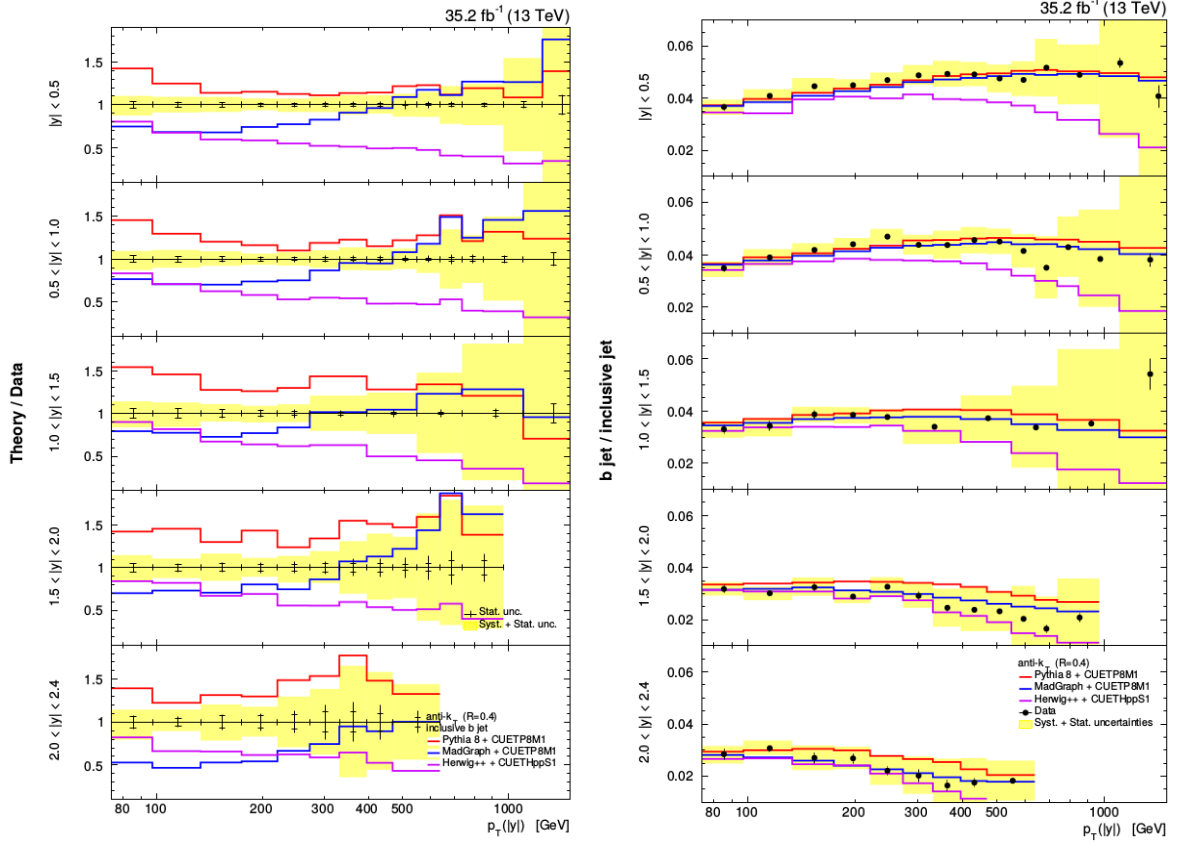
Figure 4.5. – The inclusive b-jet cross section in 13 TeV pp collisions that has been measured by CMS is compared to predictions by MadGraph [45, 46, 113], PYTHIA 8 and Herwig [49] combined with the CMS underlying-event tunes CUETP8M1 and CUETHppS1 [113] for several rapidity intervals. **Left:** The ratio of the theory predictions over the data [77]. **Right:** The ratio of the b-jet cross section over the cross section for jets of inclusive flavours [77].

5. Reconstruction of beauty jets in pp collisions at $\sqrt{s} = 13\,\text{TeV}$

In this chapter, the individual analysis steps for the measurement of beauty jets in ALICE data for pp collisions at 13 TeV are described. Details on the data sets and on the utilised MC productions are listed in Sect. 5.1. As a first analysis step, individual pp collisions – in the following called *events* – with suitable properties are selected. This event selection, together with the selection of suitable track and jet objects, provides the basis for the analysis and is discussed in Sect. 5.2. Sect. 5.3 introduces the configuration of the beauty-jet tagger which performs threshold cuts on the transverse impact-parameter significance Sd_{xy} of jet constituents and thus exploits the relatively long lifetime of the beauty hadrons decaying via the weak interaction. Inevitably, the tagging algorithm makes false positive as well as false negative decisions. This necessitates a performance correction of the raw b-tagged spectrum, that is described in Sect. 5.4. The performance-corrected spectrum is still smeared by detector effects and thus needs to be unfolded. The corresponding strategy is detailed in Sect. 5.5. The analysis is concluded by the estimation of systematic uncertainties in Sect. 5.6 and the final results are presented in Sect. 5.7. The analysis strategy described above starts from previous b-jet analyses by the ALICE collaboration (see Sect. 4) – in particular the IP method in [21, 102] but also the analysis in [93] – though it refines the corresponding strategies with respect to the performance estimation, the rejection of background from V^0 particles and the estimation of systematic uncertainties. Technical details that are beyond the scope of this thesis can be found in [86].

The final result of this thesis – apart from the refined b-jet tagging procedure – is the $p_{\text{T,ch jet}}$-differential b-jet spectrum for 13 TeV pp collisions, that is defined by the following formula

$$\frac{\mathrm{d}\sigma_{\text{ch b jet}}}{\mathrm{d}p_{\text{T,ch jet}}} = \frac{C_{\text{vtx}}}{N_{\text{ev},z_{\text{vtx}}<10\text{cm}}}\,\sigma_{\text{V0}} \cdot \text{Unfolded}\left(\frac{\mathrm{d}N_{\text{ch b-jet}}^{\text{raw}}}{\mathrm{d}p_{\text{T,ch jet}}}\,\frac{f_{\text{b}}^{\text{Tag}}}{\epsilon_{\text{b}}}\right). \tag{5.0.1}$$

Eqn. 5.0.1 introduces the basic ingredients of the analysis that can be summarised as follows:

- **raw tagged b-jet spectrum** $\frac{\mathrm{d}N_{\text{ch b-jet}}^{\text{raw}}}{\mathrm{d}p_{\text{T,ch jet}}}$:
  Charged tracks with a minimum track transverse momentum of $0.15\,\text{GeV}/c$ are clustered with the anti-k_{T} algorithm in jets with a radius of $R = 0.4$. The b-jet tagging is performed via $p_{\text{T,ch jet}}$-dependent thresholds cuts on the tracks with the largest and the second largest Sd_{xy} within jets. See Sect. 5.3 for details.

- **tagging efficiency** ϵ_{b} **and purity** $f_{\text{b}}^{\text{Tag}}$:
  The efficiency and purity of the tagging algorithm are estimated via template fits to the distributions of the jet probability $-\ln(\text{JP})$ in data[1]. This performance correction is done separately for every $p_{\text{T,ch jet}}$ bin. The fit templates are obtained from simulations with PYTHIA 8 and GEANT (PYTHIA+GEANT). See Sect. 5.4 for details.

---

[1]For simplicity, the minus sign of the observable $-\ln(\text{JP})$ will in the following be omitted as long as no quantitative statements are made. Naturally, any mentioning of $\ln(\text{JP})$ in this analysis is thus meant to refer to $-\ln(\text{JP})$

- **unfolding**:
  The final b-jet spectrum is obtained via unfolding the performance-corrected b-jet spectrum using Iterative Bayesian Unfolding [83]. The response matrix is built from PYTHIA+GEANT simulations. See Sect. 5.5 for details.

- **vertex-reconstruction efficiency C_{vtx}**:
  Only events with a vertex with a z coordinate $|z| < 10\,\text{cm}$ are utilised for the analysis. Thus, the b-jet spectrum needs to be corrected for the respective vertex reconstruction efficiency which has been determined to be $C_{\text{vtx}} = 0.952$ using the method described in [91].

- **number of accepted events $N_{\text{ev},z_{\text{vtx}}<10\text{cm}}$:**
  The b-jet spectrum is scaled by the number of events with a reconstructed vertex with $|z| < 10\,\text{cm}$.

- **reference cross section σ_{V0}:**
  The final b-jet cross section is obtained by scaling the unfolded spectrum with the cross section $\sigma_{\text{V0}} = (58.10 \pm 1.57)\,\text{mb}$ [37] for minimum bias interactions as measured by the V0 detectors of ALICE.

5.1. Data sets, Monte Carlo samples and analysis software

In this thesis, one third of the available ALICE data for pp collisions at 13 TeV is utilised. In more detail, the 2017 LHC run periods LHC17h,i,j,k,l,m,o,r with an integrated luminosity of $11\,\text{nb}^{-1}$ [37] are analysed. The analysis can be extended to take into account the data sets from 2016 and 2018.

The input from MC simulations is inevitable for various steps of data analysis in particle physics. In case of this thesis, the tuning of the b-jet tagger, its performance estimation and the unfolding of detector effects are primarily based on MC information. The main MC event generator that is used within the ALICE collaboration for the simulation of particles in the QCD vacuum is PYTHIA 8. The MC production that is utilised for this thesis has been obtained using PYTHIA 8.210 with the Monash 2013 tune [148] and the corresponding default PDF NNPDF2.3 LO [50, 70]. Hard QCD $2 \rightarrow 2$ processes have been simulated with the HardQCD:all flag of PYTHIA in 20 $\hat{p}_{\text{T}}$ bins from 5 to 235 GeV/c to assure sensible statistics also at large $p_{\text{T,ch jet}}$. Per $\hat{p}_{\text{T}}$ bin, $4 \cdot 10^7$ events have been simulated.

The *particle-level* MC information from PYTHIA is complemented by the *detector-level* information which is obtained by a propagation of the simulated particles through a model of the ALICE detector with GEANT 3 [63]. Thereby, each ALICE MC production is *anchored* to a specific LHC run in a way that the detector setup and performance during this run is mirrored by the corresponding GEANT simulation.

The MC productions are stored using the same format as the data and the MC *pseudo data* undergoes the same sequence of analysis steps as the real data. The production which is applied for this thesis is anchored to the 2017 ALICE data and is referred to as LHC18f5.

Due to the significant size of the ALICE data and MC productions of up to several hundred terabytes, the preprocessing of data and MC as well as the user analyses are executed using the computation power of the Worldwide LHC Computing Grid (WLCG). The corresponding services, like workload and data management, are provided by a set of Grid Middleware Services called AliEn [8]. After the analysis on the grid, the data has been compressed to a size of several gigabytes in case of this thesis and can be downloaded for further analysis on local PCs.

The C++ framework ROOT [62] is nowadays widely used for the analysis and the graphical visualisation of data in the field of particle physics. It was developed for data analysis at CERN and offers a large selection of tools for statistical analysis. In this thesis, ROOT is applied for the

Table 5.1. – Criteria for the primary-vertex selection. Here, z_{SPD} is the z-coordinate of the vertex that is reconstructed from SPD information whereby z_{tracks} is the z-coordinate that has been reconstructed from tracks.

vertex property	selection criterion
number of contributors	≥ 1
z-coordinate	$\lvert z\rvert < 10\,\mathrm{cm}$
$\lvert z_{\mathrm{SPD}} - z_{\mathrm{tracks}}\rvert$	$< 0.5\,\mathrm{cm}$
resolution for vertices	< 0.25
with reconstructed z-coordinate only	

visualisation of results but also as an essential ingredient for several analysis steps: the RooFit framework [157] is utilised for template-fitting, as discussed in Sect. 5.3, and the RooUnfold package [28] is used to unfold detector effects, as described in Sect. 5.5.

In ALICE, the first basic analysis steps like cluster-, vertex-, and track finding in the raw data are performed offline and centrally for all users. The corresponding functionality is implemented in AliRoot [33], which is an object-oriented framework based on ROOT. The data runs through several so-called *reconstruction passes*, in which the information from the reconstruction is optimised. After each pass, the properties of the events are stored as *Event Summary Data* (ESD). These can, in principle, serve as input to the analysis by endusers. However in general, the ESDs are further compressed into *Analysis Object Data* (AOD) for which information is condensed and which are thus much lighter. This analysis uses AOD of reconstruction pass1 for the 2017 ALICE data.

The more specific analyses by the endusers are combined in the AliPhysics framework [32]. Within this framework, common functionalities like predefined cuts for the event selection are available as ready-to-use while specialised components are implemented in a modular way by the endusers. This thesis also makes use of the ALICE jet-finding framework [34] that adjusts the functionality of Fastjet 3.2.1 [67, 69] in a way that is customised for the identification and analysis of jet objects within ALICE.

5.2. Event, track and jet selection

5.2.1. Event selection

This analysis uses events that have been selected by the V0AND minimum bias trigger (see Sect. 3.2.1 for definition) of ALICE. The respective minimum bias cross section for the 2017 ALICE data is $\sigma_{\mathrm{V0}} = (58.10 \pm 1.57)\,\mathrm{mb}$ [37]. Only events with a high-quality primary vertex have been analysed. In particular, data from the SPD are required for the vertex reconstruction and the vertex needs to be situated within $\lvert z\rvert < 10\,\mathrm{cm}$ from the nominal interaction point. The latter necessitates the correction by the vertex reconstruction efficiency, as it is described at the beginning of this chapter. The full list of vertex selection criteria is provided in Tab. 5.1.

It may happen that more than one pp collision take place in the ALICE detector within the readout time of the subdetectors – an effect that is called pile-up. Pile-up events are rejected based on multiple vertices that have either been reconstructed by the SPD or from tracking information. In addition, a selection on the correlation of the number of SPD clusters in dependence on the number of SPD tracklets is applied.

After the event selection, about $6 \cdot 10^8$ events remain for the analysis.

Table 5.2. – Selection criteria that are applied to tracks used for the estimation of the b-jet tagging decision. A kink candidate is a track with a distinct kink as it is e.g. produced by the decay $K^+ \to \mu^+ + \nu_\mu$. The decay length is the distance between the primary-vertex position and the position of the track at the DCA to the jet.

track property	selection criterion
track p_T	$0.15\,\mathrm{GeV}/c$
track η	$|\eta| < 0.9$
ITS hits	4
SPD hits	2
number of TPC clusters	> 100
ITS refit	$\checkmark$
TPC refit	$\checkmark$
no kink candidates	$\checkmark$
track χ^2/NDF	< 5
decay length	$< 5\,\mathrm{cm}$
d_0	$< 1\,\mathrm{cm}$
d_z	$< 2\,\mathrm{cm}$
$\mathrm{DCA_{jet\text{-}track}}$	$< 0.07\,\mathrm{cm}$

5.2.2. Track selection

Two categories of track selection criteria are applied for this analysis. A standard selection is deployed for tracks that are used for the jet finding. In this case, charged tracks with $p_T > 0.15\,\mathrm{GeV}/c$ and a pseudorapidity of $|\eta| < 0.9$ are utilised. To overcome an inhomogeneous track distribution as a function of η and φ due to inefficiencies in the SPD, a combination of "global tracks" (high-quality tracks with hits in the SPD) and "complementary tracks" (tracks without SPD hits) is used. This method is called hybrid tracking in the ALICE collaboration [13]. The identification of beauty jets is performed based on the track properties of tracks within jets. In order to obtain reliable information from these tracks, it is essential that they have been reconstructed with excellent quality. This is why, after the jet finding, only tracks passing strict selection criteria are allowed for the estimation of the tagging decision. In particular, selections on the impact parameter in the xy-plane ($d_0 < 1\,\mathrm{cm}$) as well as in z direction ($d_z < 2\,\mathrm{cm}$) are applied to reject background from decay tracks of long-lived V^0 particles. The distance of closest approach (DCA) between the jet and its track constituents $\mathrm{DCA_{jet\text{-}track}}$ is asked to be smaller than $0.07\,\mathrm{cm}$ for the same reason. The full list of selection criteria for tracks that are used as input for the b-jet tagging decision are listed in Tab. 5.2.

5.2.3. Jet selection

Jets are reconstructed from charged-particle tracks using the anti-k_T algorithm with E-scheme recombination [68] as implemented in Fastjet. They are required to have a radius $R = 0.4$ and to be within the fiducial acceptance of the TPC, which entails a pseudorapidity of $|\eta_{\mathrm{Jet}}| < 0.5$. The jet area is asked to be larger than $60\,\%$ of the expected area πR^2 and the pion mass is assumed for the track constituents. The final b-jet spectrum will be reported for a transverse jet momentum of $10\,\mathrm{GeV}/c$ to $100\,\mathrm{GeV}/c$

For jets in the MC simulations, the flavour of a jet is defined based on the most massive parton within the jet cone. Thereby, a parton is assumed to be within the jet cone if $\sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2} < R$ is valid, whereby $\Delta\eta$ and $\Delta\varphi$ are the angular differences in η and φ of the quark momentum and the jet axis. Since this definition does not include information about the process the quark originates from, it is not possible to separate gluon jets from quark jets.

5.3. Beauty-jet tagging

In this thesis, the selection of beauty jets is performed based on the signed transverse impact parameter significance Sd_{xy} of tracks within jets as it was introduced in Sect. 4. The Track Counting Algorithm that is utilised in this respect will be described in the next subsection. The identification of b jets as well as the estimation of the tagging performance rely on the capability of the MC simulations to capture the Sd_{xy} distributions in data. This is why, Sect. 5.3.2 discusses the difference between the corresponding distributions in data and in MC in the context of this issue.

5.3.1. The Track Counting algorithm

The Track Counting Algorithm has already been applied for b-jet tagging in other analyses, e.g. by CMS [75] or ALICE [21]. It sorts the tracks within a jet in descending order of their Sd_{xy} values. Thereby, the tracks with the largest Sd_{xy} shall be called *N1-type tracks* and the ones with smaller impact parameters *Ni-type tracks* with i giving the position in the Sd_{xy} ranking within a single jet. Based on this ranking of tracks, threshold selections can be applied to identify beauty jets. In this analysis, it was decided to impose simultaneous cuts on the Sd_{xy} of N1- and N2-type tracks as this setting was found to maximise the product of the tagging purity and efficiency in MC simulations.

The threshold values have been determined separately for N1- and N2-type tracks and also separately for every $p_{\text{T,ch jet}}$ bin of the final b-jet spectrum. With this $p_{\text{T,ch jet}}$-dependence, it is intended to achieve b-jet tagging efficiencies and purities that are roughly constant over the considered $p_{\text{T,ch jet}}$ range. To estimate the threshold values, the fit function provided in Sect. A.1 is used to fit the positive side of the probability distributions (p.d.f.s) for the Sd_{xy} of the individual jet flavours, that is obtained from PYTHIA+GEANT simulations. Exemplary fits to Sd_{xy} distributions of N1-type tracks within light-flavour, c and b jets are shown for $30 < p_{\text{T,ch jet}} < 40 \,\text{GeV}/c$ in Fig. 5.1.

The resulting fit functions can be identified with the distribution of the conditional probability $P(Sd_{xy}|i)$ that a jet of flavour $i \in [\text{uds,c,b}]$ contains tracks with a specific Sd_{xy}. Exploiting Bayes Theorem [59], the probability $P(Sd_{xy}|\text{b})$ can be related to the conditional probability $P(\text{b}|Sd_{xy})$ that a track with a certain Sd_{xy} is a b-jet constituent. The relation that has been utilised in this thesis was inspired by [116] and reads as follows

$$p \equiv P(\text{b}|Sd_{xy}) = \frac{P(Sd_{xy}|\text{b})P(\text{b})}{P(Sd_{xy}|\text{uds})P(\text{uds}) + P(Sd_{xy}|\text{c})P(\text{c}) + P(Sd_{xy}|\text{b})P(\text{b})}.$$

Thereby, $P(i)$ is the respective abundance of the jet flavour i in the investigated $p_{\text{T,ch jet}}$ bin and the conditional probability $P(\text{b}|Sd_{xy})$ shall be referred to as the *separation power p* in the following.

The Sd_{xy} distributions for every $p_{\text{T,ch jet}}$ bin are now analysed binwise with respect to their separation power, going from large to low Sd_{xy} values. The Sd_{xy} value of the largest-Sd_{xy} bin for which the p value exceeds a predetermined value is chosen as respective threshold value. Every working point of the b-jet tagger is thus characterised by a p value based on which the corresponding threshold values are determined.

The method described above for the estimation of the tagger working point in principle enables the flexible adjustment of the tagger setting to the individual needs of an analysis with respect to tagging purity and efficiency. However, it has been found that the method that is used for the performance estimation in this thesis puts technical constraints on the choice of the threshold values – this is discussed in more detail in Sect. 5.4.1. Investigations have shown that the tagger setting $p = 0.10$ offers a reasonable compromise between the tagging performance and the uncertainties that arise from the performance estimation. Consequently, this setting is used
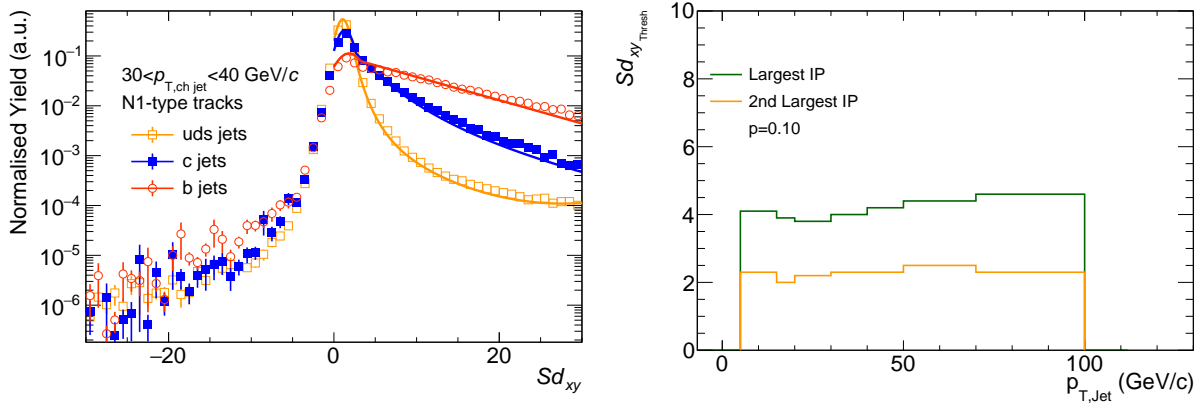
Figure 5.1. – **Left:** Exemplary fits to the Sd_{xy} distributions for N1-type tracks in light-flavour, c and b jets from MC simulations as they are used for the determination of the b-tagging threshold values. The distributions are normalised to unity. **Right:** Threshold values for N1- and N2-type tracks for the default setting of the b-jet tagger with a separation power $p = 0.10$.

as the default working point of the b-jet tagger. The corresponding threshold values for N1- and N2-type tracks are shown in Fig. 5.1. One can see that the $p_{T,\text{ch jet}}$ dependence for the threshold values of the N2-type tracks is negligible whereas those for N1-type tracks are slightly increasing with $p_{T,\text{ch jet}}$.

5.3.2. Impact-parameter distributions for MC simulations and data

A reasonable description of the Sd_{xy} distributions in data by the MC simulations is essential for the method that is used for the performance estimation in Sect. 5.4. As it can be seen in Fig. 5.2 for three different intervals of $p_{T,\text{ch jet}}$, the central peaks of the Sd_{xy} distributions for tracks within jets of inclusive flavours are consistent between data and MC. However, the Sd_{xy} distributions in data are wider in the tails than the distributions in MC, which leads to a significant underestimation of the data for large $|Sd_{xy}|$. For the $N1$- and $N2$-type tracks, a similar behaviour can be observed as shown in Sect. A.2.

5.4. Performance estimation

To minimise the reliance on MC information, the performance of the b-jet tagger is investigated by a data-driven procedure using template fits to distributions of the jet probability ln(JP). Thereby, the tagging efficiency and purity is obtained from separate fits to the data sample before (denoted by *untagged*) and after the b-jet tagging (denoted by *b-tagged*) for every $p_{T,\text{ch jet}}$-bin of the final spectrum. The fit templates are the ln(JP) distributions for light-flavour, c and b jets from PYTHIA+GEANT simulations. The basic concept of the template fitting procedure is described in Sect. 5.4.2. Sect. 5.4.3 refines the basic approach by technical details that improve the stability and reliability of the template fit results. Finally, the results of the performance estimation are provided in Sect. 5.4.4.
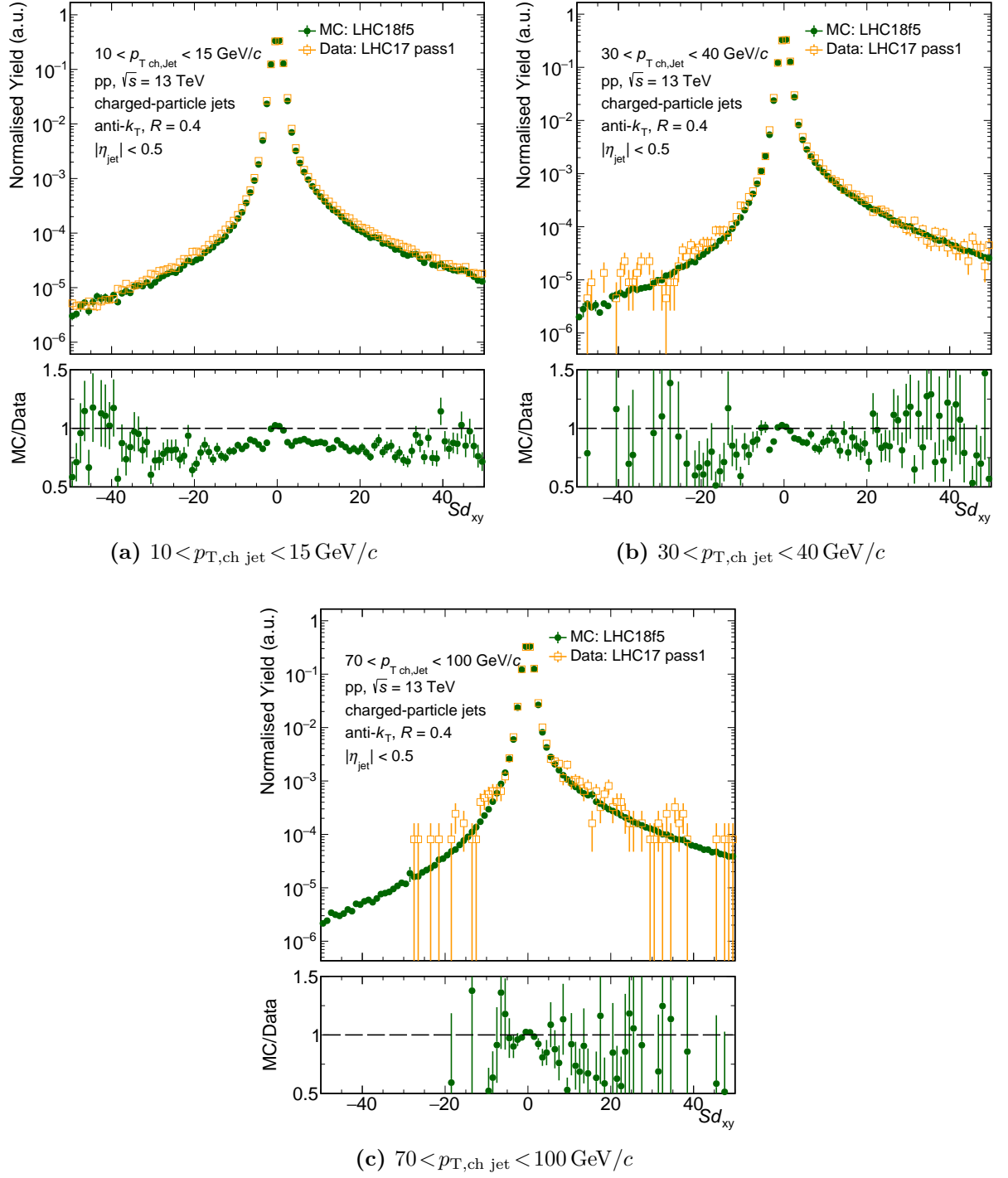
(a) $10 < p_{\mathrm{T,ch\ jet}} < 15\,\mathrm{GeV}/c$

(b) $30 < p_{\mathrm{T,ch\ jet}} < 40\,\mathrm{GeV}/c$

(c) $70 < p_{\mathrm{T,ch\ jet}} < 100\,\mathrm{GeV}/c$

Figure 5.2. – Comparison of the Sd_{xy} distributions for tracks within inclusive jets between data and MC at low, medium and high $p_{\mathrm{T,ch\ jet}}$. The distributions are normalised to unity.

5.4.1. The Jet Probability ln(JP)

The jet probability ln(JP) provides a measure for the probability that a jet originates from a secondary vertex: the larger $-\ln(\text{JP})$, the larger the probability for the existence of a secondary vertex within the jet cone. The observable is defined via the *track probability*

$$P_{\text{tr}}(I_0) = \frac{\int\limits_{-\infty}^{-|I_0|} R(I)\mathrm{d}I}{\int\limits_{-\infty}^{0} R(I)\mathrm{d}I} \tag{5.4.1}$$

which quantifies the probability that a track is coming from the primary vertex. Thereby, $R(I)$ denotes the so-called *resolution function* that is defined in more detail below. I_0 is the impact parameter significance for the track that is under investigation. For the definition of the jet probability, the track probability $P_{\text{tr},k}$ of the N_{tr} tracks within one jet are combined by a product as

$$\text{JP} = \Pi \cdot \sum_{k=0}^{N_{\text{tr}}-1} \frac{(-\ln\Pi)^k}{k!} \quad \text{with} \quad \prod = \prod_{k=1}^{N_{\text{tr}}} P_{\text{tr},k}. \tag{5.4.2}$$

The logarithm of JP provides a better separation power, thus, the observable $-\ln(\text{JP})$ is often utilised for the b-jet identification [21, 141]. Since b hadrons decay at a secondary vertex that is well displaced from the primary vertex (see Sect. 4), $-\ln(\text{JP})$ distributions are wider for b jets than for light-flavour or c jets as it can be seen in Fig. 5.3.

Only tracks with $Sd_{xy} > 0$ are considered for the estimation of ln(JP) in this thesis, which is why not for every jet a corresponding ln(JP) can be defined.

Definition of the resolution function   The resolution function $R(I)$ that enters Eqn. 5.4.1 is the Sd_{xy} distribution for tracks within jets of inclusive flavours for $Sd_{xy} < 0$. Since the resolution of the Sd_{xy} differs in data and MC (see Sect. 5.3.2), the results for the performance estimation will be biased if this difference is not taken into account. Going into more detail, these biases can arise for two reasons:

1. a discrepancy of the $R(I)$ that is utilised for the ln(JP) estimation in MC and data as well as

2. a discrepancy of the $R(I)$ that is used for the estimation of the ln(JP) with respect to the intrinsic resolution of the sample either in data or MC.

In previous analyses [21, 141], the resolution function $R_{\text{data}}(I)$ from data has been used to calculate the ln(JP) in data and the resolution function $R_{\text{MC}}(I)$ from MC has been applied to obtain the ln(JP) in MC. As in this analysis, the statistics of the data is limited at large $p_{\text{T,ch jet}}$, the $R_{\text{MC}}(I)$ is used to determine the ln(JP) both in data and in MC. Deviations of the estimated performance observables that arise from biases 1) and 2) are investigated as *shape uncertainties* in Sect. 5.6.1.

The Sd_{xy} distributions depend on jet and track properties. This is why $R_{\text{MC}}(I)$ has been determined separately for

- four $p_{\text{T,ch jet}}$ bins (respective binning: $10, 15, 20, 40, 100\,\text{GeV}/c$),

- tracks with momentum $p_{\text{T,tr}} \leq 2\,\text{GeV}/c$ and $p_{\text{T,tr}} > 2\,\text{GeV}/c$ and

- N1,N2,N3,N4 and Ni-type tracks with $i \geq 5$, in the following denoted as N5.

The corresponding distributions for jets of inclusive flavours have been fit with the fit function defined in Sect. A.1 for $Sd_{xy} < 0$ to obtain smooth parametrisations. The Sd_{xy} distributions for
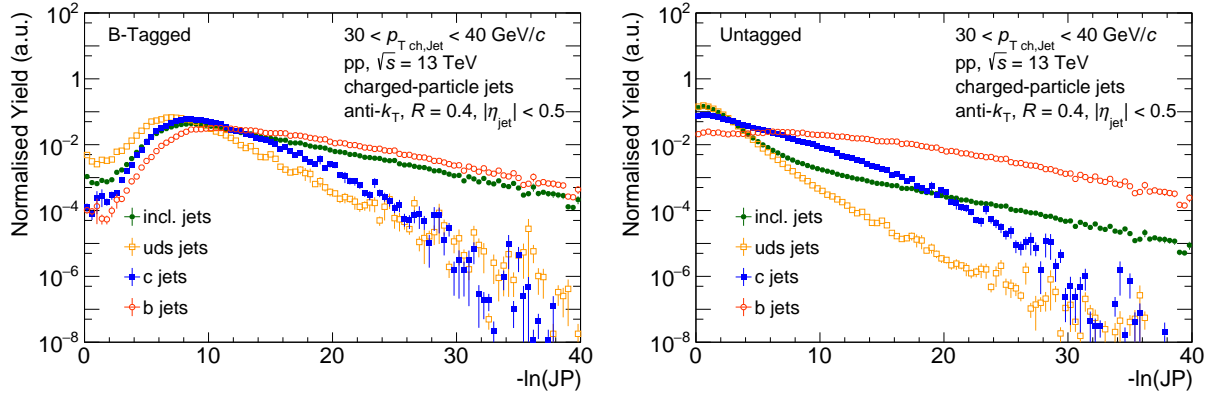
Figure 5.3. – PYTHIA+GEANT simulations of the ln(JP) distributions for different jet flavours and $30 < p_{\mathrm{T,ch\ jet}} < 40\,\mathrm{GeV}/c$ for b-tagged (left) and b-untagged (right) jets. Each distribution is normalised to unity.

tracks within jets of inclusive flavours with $20 < p_{\mathrm{T,ch\ jet}} < 40\,\mathrm{GeV}/c$ are shown exemplarily in Fig. 5.4 together with the respective fit functions.

Finally, the ln(JP) for each jet has been calculated from integrating the corresponding $R_{\mathrm{MC}}(I)$ as defined in Eqn. 5.4.1 in dependence of $p_{\mathrm{T,ch\ jet}}$ and the properties of the jet constituents.

Resulting ln(JP) distributions in data and MC simulations Fig. 5.3 shows the ln(JP) distributions for different jet flavours and b-tagged as well as b-untagged jets with $30 < p_{\mathrm{T,ch\ jet}} < 40\,\mathrm{GeV}/c$ from PYTHIA+GEANT simulations. As it was stated already before, b jets exhibit wider ln(JP) distributions than light-flavour or c jets. The ln(JP) distributions for the b-tagged sample show a suppression at small $-\ln(\mathrm{JP})$ due to the Sd_{xy} threshold selection of the tagging criterion. This leads to distributions for the individual jet flavours which show increasing similarity with stronger selection criteria. Since this similarity evokes stability issues for the corresponding ln(JP) template fits, the choice of the tagger working point is limited to lower threshold values for this method of performance estimation.

Fig. 5.5 compares the ln(JP) distributions for jets of inclusive flavours in untagged data and PYTHIA+GEANT simulations for low, medium and large $p_{\mathrm{T,ch\ jet}}$. One can see that for low $p_{\mathrm{T,ch\ jet}}$, the MC distributions underestimate the distributions in data for $3 \lesssim -\ln(\mathrm{JP}) \lesssim 15$ by up to 20 %. This effect is not visible for the b-tagged distribution in Fig. 5.6. At high $p_{\mathrm{T,ch\ jet}}$, the MC simulations are consistent with the tagged and the untagged data within uncertainties.

5.4.2. Basic concept of the template-fitting procedure

The ln(JP) distributions for the different jet flavours are normalised to unity and interpreted as p.d.f.s. The fit model for the template-fitting procedure is then given by the weighted sum of three p.d.f.s $\mathcal{P}_i$ for light-flavour, c and b jets (i.e. $i \in \mathrm{lf, c, b}$). In particular, the predicted total number of jets $N_{\mathrm{lnJP}}(\ln(\mathrm{JP}))$ with well defined ln(JP) can be formulated as

$$N_{\mathrm{lnJP}}(\ln(\mathrm{JP})) = N_{\mathrm{lnJP}}^{\mathrm{lf}}\mathcal{P}_{\mathrm{lf}}(\ln(\mathrm{JP})) + N_{\mathrm{lnJP}}^{\mathrm{c}}\mathcal{P}_{\mathrm{c}}(\ln(\mathrm{JP})) + N_{\mathrm{lnJP}}^{\mathrm{b}}\mathcal{P}_{\mathrm{b}}(\ln(\mathrm{JP})) \qquad (5.4.3)$$

for untagged jets and in the same way for b-tagged jets. The free fit parameters are the number of jets N_{lnJP}^i of a specific flavour. The fits are performed using the functionality of the RooFit package [157] for Extended Maximum Likelihood fits [52, 110, 155]. Maximum-likelihood fits outperform $\mathcal{X}^2$ fits in environments of small statistics for which no Gaussian uncertainties can be assumed. They are thus well suited for fitting the ln(JP) distributions obtained in this thesis also for large jet momenta. Extended maximum likelihood fits differ from regular maximum likelihood fits in a way that the predicted total number of events is allowed to fluctuate around
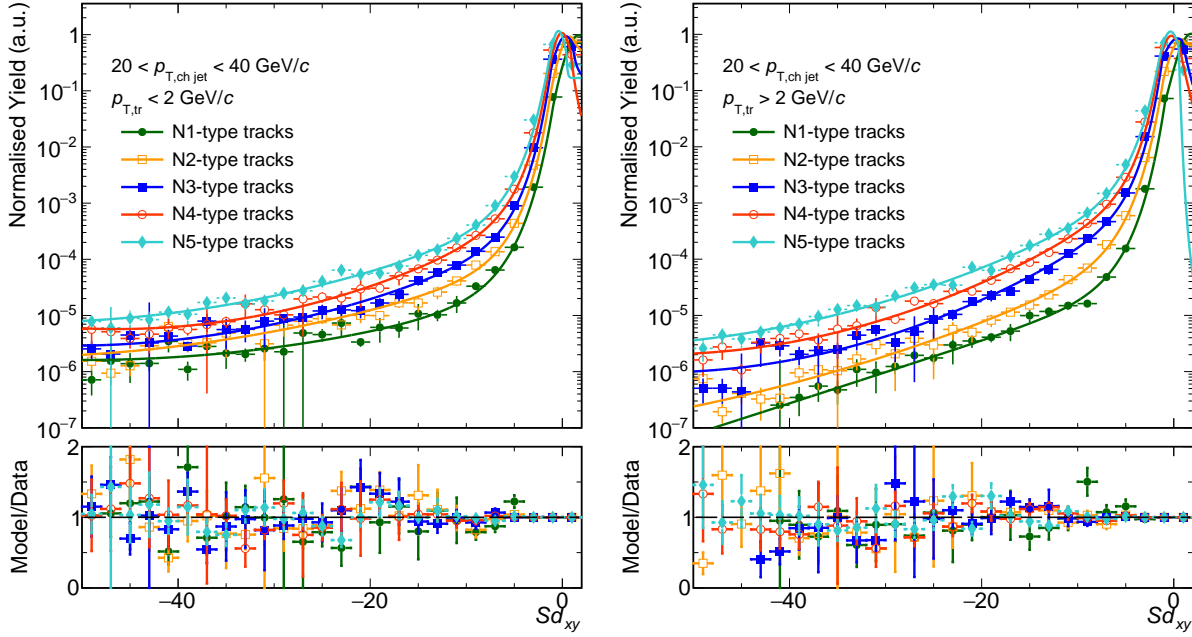
Figure 5.4. – Exemplary fits to the Sd_{xy} distributions for Ni-type tracks with $p_{\mathrm{T,tr}} < 2\,\mathrm{GeV}/c$ (left) and $p_{\mathrm{T,tr}} > 2\,\mathrm{GeV}/c$ (right) within jets of inclusive flavours as they are used for the determination of the jet probability $\ln(\mathrm{JP})$. The distributions have been obtained by PYTHIA+GEANT simulations and are normalised to unity.

the measured events according to a Poisson distribution. This solves complications with respect to the overall normalisation of the fit model, which arise from its composition of three p.d.f.s and which make it a *type B* problem following the nomenclature in [52].

The efficiency and the purity of the b-jet tagger are determined from the template-fit results for the b-jet fraction in the untagged and the tagged data sample. Thereby, the b-jet fraction in the untagged data sample is calculated as

$$f_{\mathrm{b}} = N^{\mathrm{b}}_{\ln\mathrm{JP}}/N_{\ln\mathrm{JP}} \quad \text{with} \quad N_{\ln\mathrm{JP}} = \int N_{\ln\mathrm{JP}}(\ln(\mathrm{JP})) \tag{5.4.4}$$

and the number of b jets $N^{b}_{\ln\mathrm{JP}}$ with well defined $\ln(\mathrm{JP})$. The b-jet fraction in the b-tagged data sample $f^{\mathrm{Tag}}_{\mathrm{b}}$ corresponds to the tagging purity and is defined equivalently. The efficiency ϵ_{b} of the b-jet tagger is obtained as

$$\epsilon_{\mathrm{b}} = C_{\mathrm{b}} \frac{f^{\mathrm{Tag}}_{\mathrm{b}} N^{\mathrm{Tag}}_{\ln\mathrm{JP}}}{f_{\mathrm{b}} N_{\ln\mathrm{JP}}} \tag{5.4.5}$$

with the number of jets $N^{\mathrm{Tag}}_{\ln\mathrm{JP}}$ with well-defined $\ln(\mathrm{JP})$ (see Sect. 5.4.1) in the tagged sample and the number of jets $N_{\ln\mathrm{JP}}$ in the untagged sample [21]. C_{b} is the fraction of b-jets with well-defined $\ln(\mathrm{JP})$, which is determined from MC simulations. The behaviour of C_{b} in dependence of $p_{\mathrm{T,\,jet}}$ is shown in Fig. 5.7: it is increasing from about 93 % to 97 % for a transverse momentum of 10 to 20 GeV/c, from where it stays constant up to 100 GeV/c.

For the estimation of the statistical uncertainties, $N_{\ln\mathrm{JP}}$ and $N^{\mathrm{Tag}}_{\ln\mathrm{JP}}$ are considered to be fully correlated, while all other pairings ob observables are assumed to be fully uncorrelated. For more information regarding parameter correlations, it is referred to Appendix B.1.

The default fit range is defined as $0 < -\ln(\mathrm{JP}) < 28$. Exemplary template fits for the default fit range are shown for the b-tagged and the untagged data sample and $30 < p_{\mathrm{T,ch\ jet}} < 40\,\mathrm{GeV}/c$ in Fig. 5.8. The respective results for all other $p_{\mathrm{T,ch\ jet}}$ bins are provided in Appendix B.2. The
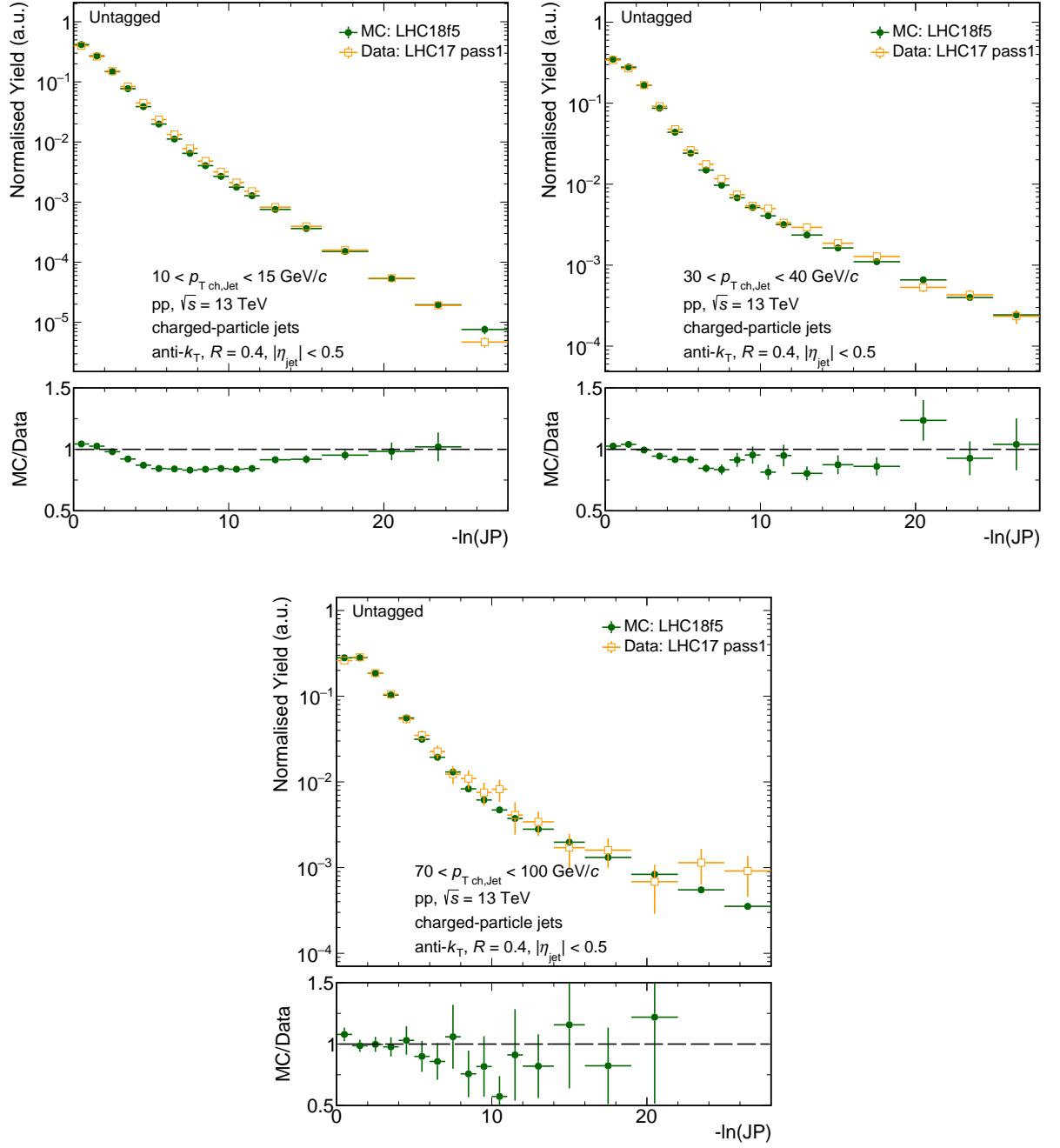
Figure 5.5. – Comparison of the ln(JP) distributions for jets of inclusive flavours between PYTHIA+GEANT simulations and data for low $(10 < p_{T,ch\ jet} < 15\ GeV/c)$, medium $(30 < p_{T,ch\ jet} < 40\ GeV/c)$ and large $p_{T,ch\ jet}$ $(70 < p_{T,ch\ jet} < 100\ GeV/c)$. The distributions are shown for **untagged jets** and are normalised to unity.
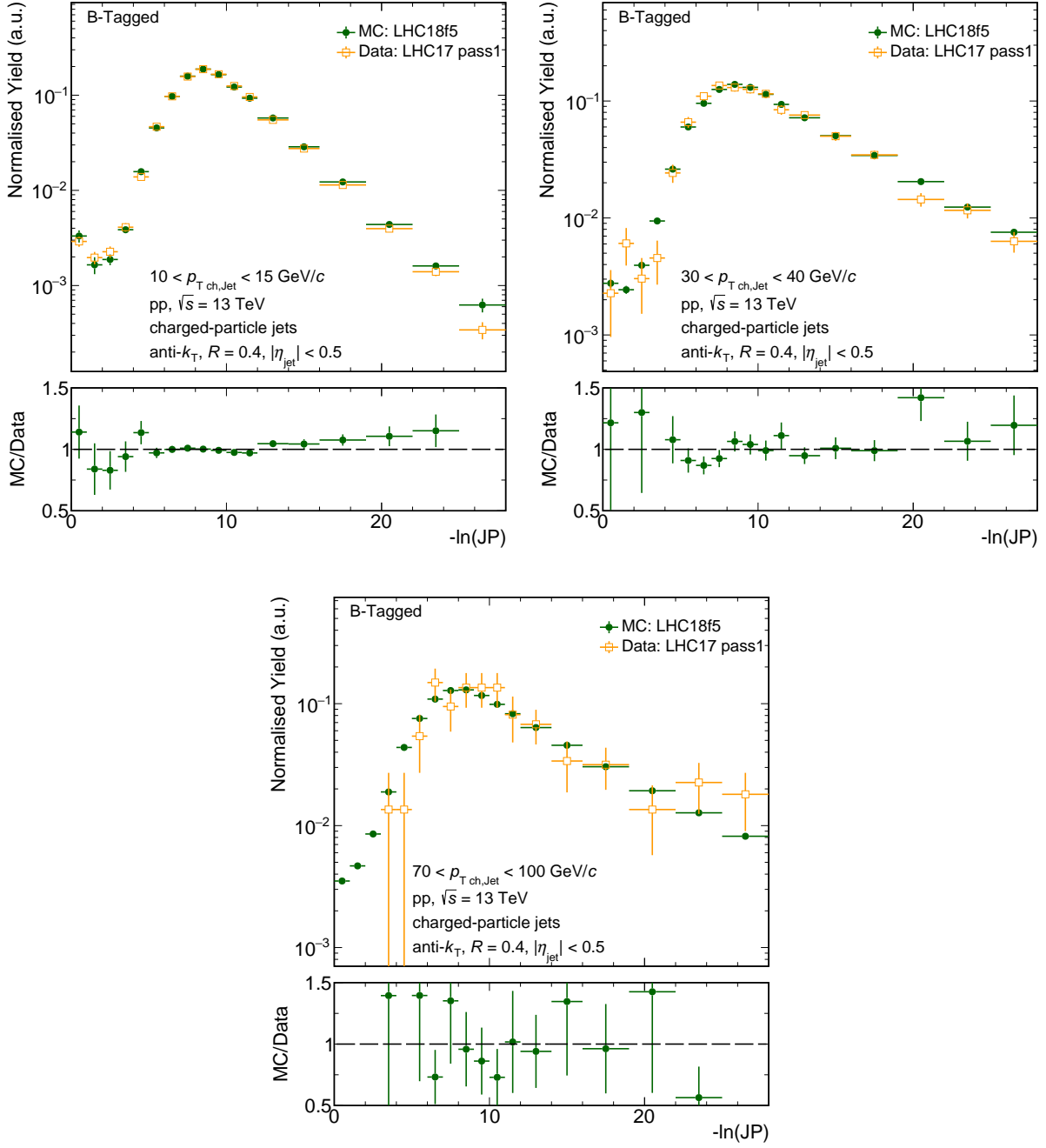
Figure 5.6. – Comparison of the ln(JP) distributions for jets of inclusive flavours between PYTHIA+GEANT simulations and data for low ($10 < p_{\mathrm{T,ch\ jet}} < 15\,\mathrm{GeV}/c$), medium ($30 < p_{\mathrm{T,ch\ jet}} < 40\,\mathrm{GeV}/c$) and large $p_{\mathrm{T,ch\ jet}}$ ($70 < p_{\mathrm{T,ch\ jet}} < 100\,\mathrm{GeV}/c$). The distributions are shown for **b-tagged jets** and are normalised to unity.
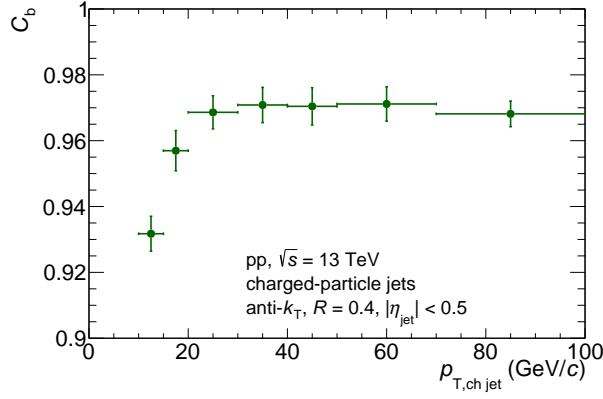
Figure 5.7. – The fraction C_b of b-jets with well-defined $\ln(\mathrm{JP})$ in dependence of $p_\mathrm{T,ch\ jet}$.
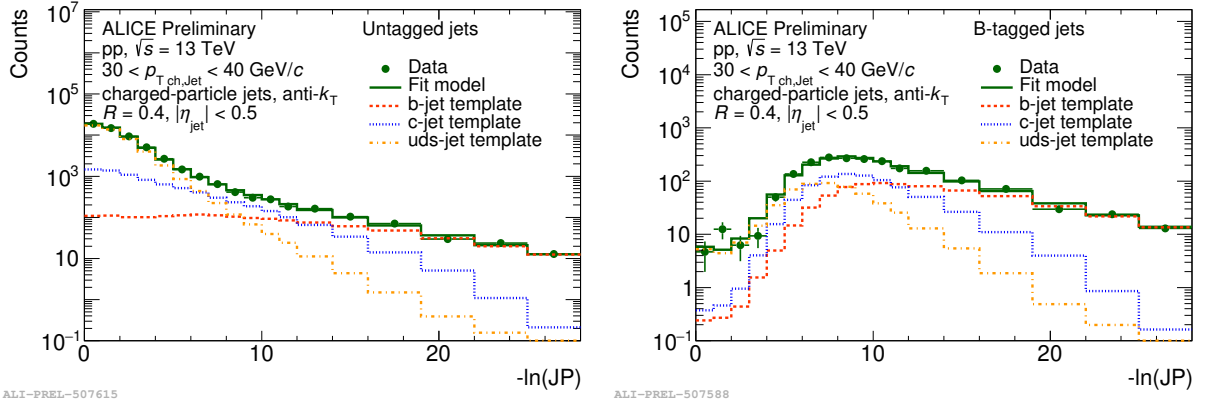

Figure 5.8. – The fit templates, the fit model and the data for fits to the untagged (left) and the b-tagged (right) data for the default fit range and $30 < p_\mathrm{T,ch\ jet} < 40\,\mathrm{GeV}/c$. The fit templates for the individual jet flavours are scaled with respect to their relative contribution to the total fit model.

performance-corrected b-jet spectrum is obtained by multiplying the raw b-tagged spectrum with the purity $f_\mathrm{b}^\mathrm{Tag}$ and dividing by the efficiency ϵ_b, as defined in Eqn. 5.0.1.

5.4.3. Improving the stability and reliability of the template-fit results

Results from fits with composite fit models can suffer from different issues with respect to stability and reliability. Complications can, for example, arise if the individual p.d.f.s that form the fit model exhibit similar shapes. Corresponding issues have been experienced for $\ln(\mathrm{JP})$ template fits at low jet momentum in [21], where they entailed a restriction of the $p_\mathrm{T,ch\ jet}$ range over which the final b-jet cross section could be measured. In a case like this, the fit is underdetermined and the minimisation procedure terminates arbitrarily at different local minima. This can lead to fit results that strongly fluctuate with small changes of the data distribution and is aggravated by fits to data with low statistics. In this context, the accurate estimation of uncertainties that capture the variability of the fit results is essential. At the same time, fit templates obtained from MC simulations might not describe the data sufficiently well and thus invoke biased fit results or strong dependencies on the choice of the fit range. Finally, mistakes in the implementation of the fitting procedure can be present – an avoidable but fundamental source for deficient results. In this section, the method for the performance estimation will be examined with respect to the deficiencies listed above. Corresponding issues will be addressed with the objective of developing a template-fitting procedure that provides reliable and robust results.
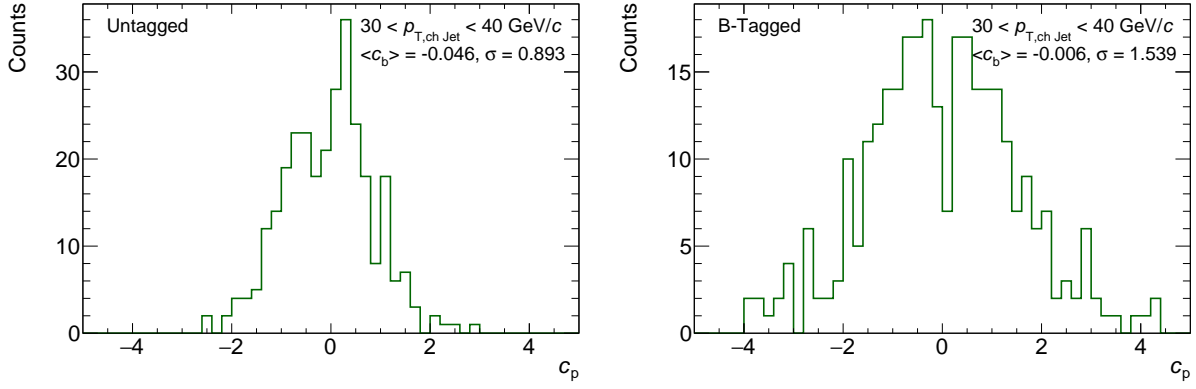
Figure 5.9. – Pull distributions from MC toy studies for 300 fits to untagged pseudo data (left) as well as b-tagged pseudo data (right) with $30 < p_{\mathrm{T,ch\ jet}} < 40\,\mathrm{GeV}/c$. The mean $\langle c_p \rangle$ and the standard deviation σ of the pull distributions are provided in the figure.

Estimation of statistical uncertainties Inspired by [156], the reliability of the statistical uncertainties that are provided by RooFit for the fit parameters is assessed by *MC toy simulations*. Thereby, 300 *toys* are obtained from the MC pseudo data. For the generation of each toy, the content of every bin of the original pseudo data is varied according to a Gaussian distribution within the statistical uncertainties. Every toy distribution is then fit with the original fit templates. The stability of this ensemble of fits is measured using the *pull distribution*

$$c_p = \frac{f_b^{\mathrm{Fit}} - f_b^{\mathrm{Gen}}}{\sigma_b^{\mathrm{Fit}}}. \tag{5.4.6}$$

In Eqn. 5.4.6, f_b^{Fit} is the b-jet fraction that is obtained from the fits to MC pseudo data, σ_b^{Fit} is the corresponding uncertainty that is provided by RooFit and f_b^{Gen} is the true MC b-jet fraction. The results for the pull distributions from fits to the b-tagged and untagged MC pseudo data for $30 < p_{\mathrm{T,ch\ jet}} < 40\,\mathrm{GeV}/c$ are shown in Fig. 5.9. The distributions are well centered around zero, which proves that the fits are unbiased. However, the standard deviations differ from one. In particular, the uncertainties are underestimated (overestimated) for fits to the b-tagged (untagged) pseudo data.

To assure that the variability of the fit results with respect to the statistical fluctuations is captured by the statistical uncertainties, the latter are estimated by *data toy studies* in this analysis. This means that, as for the MC toy studies, 300 toys are obtained from the original data. Since the real data is Poisson distributed, the toys are generated by varying the bin contents of the data according to a Poisson distribution. The distribution of the b-jet fractions for the fits to every individual toy is approximately Gaussian, as it can be recognised from Fig. 5.10. Therefore, the standard deviation and the mean of the distribution are adopted as the statistical uncertainty and the central value for the b-jet fraction in one particular $p_{\mathrm{T,ch\ jet}}$ bin. In this sense, whenever it is referred to the b-jet fraction f_b in the following, the mean of the individual fit results will be provided (i.e. $f_b \equiv \langle f_b \rangle$).

The impact of the statistical uncertainties of the fit templates on the fit results is estimated as a systematic uncertainty in Sect. 5.6.3.

Averaging results from different fit ranges It has been observed, that, especially at small $p_{\mathrm{T,ch\ jet}}$, the fit results for fits to the untagged data sample react very sensible to a change of the lower fit boundary from $-\ln(\mathrm{JP})_{\mathrm{min}} = 0$ to larger values. In Fig. 5.11, the results for f_b^{Tag} and
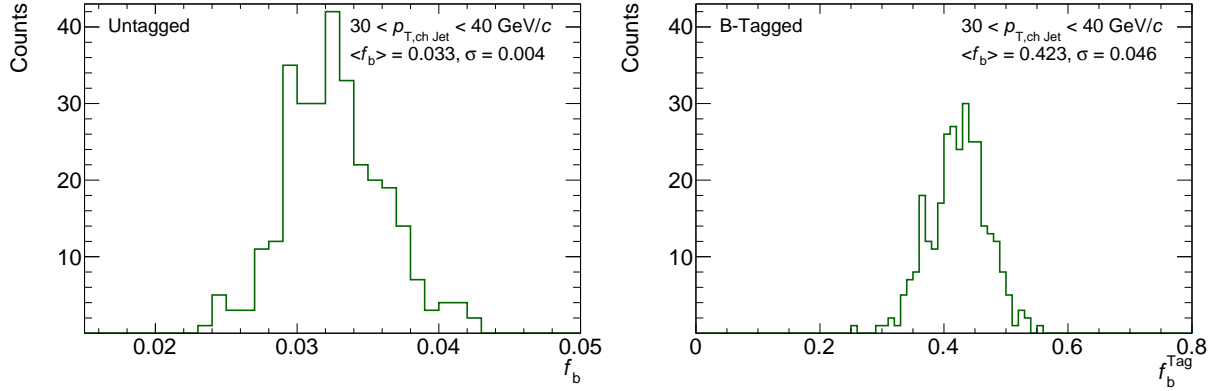
Figure 5.10. – The distributions for the b-jet fractions in the untagged (left) and the b-tagged (right) data sample for 300 template fits and $30 < p_{\mathrm{T,ch\ jet}} < 40\,\mathrm{GeV}/c$. The corresponding mean $\langle f_{\mathrm{b}} \rangle$ and the standard deviation σ of the distributions are provided in the figures.

f_{b} for all seven $p_{\mathrm{T,ch\ jet}}$ bins are shown as a function of $-\ln(\mathrm{JP})_{\mathrm{min}}$. Thereby, the results $f_{\mathrm{b,lim}}$ for the b-jet fraction determined from fits over the limited fit range have been propagated to the results for the default fit range via

$$f_{\mathrm{b}} = f_{\mathrm{b,lim}} \cdot \frac{I_{\mathrm{All,lim}}/I_{\mathrm{All}}}{I_{\mathrm{b,lim}}/I_{\mathrm{b}}}. \qquad (5.4.7)$$

Here, I_{All} and $I_{\mathrm{All,lim}}$ are the integrals of the Sd_{xy} distribution for jets of inclusive flavours for the default and the limited $\ln(\mathrm{JP})$ range and I_{b} and $I_{\mathrm{b,lim}}$ are the respective integrals for the b-jet distributions. All integrals of Eqn. 5.4.7 are determined from PYTHIA+GEANT simulations and the propagation is performed in the same manner for the b-jet fraction in the tagged data sample. One can see in Fig. 5.11, that for $10 < p_{\mathrm{T,ch\ jet}} < 20\,\mathrm{GeV}/c$ the b-jet fraction increases significantly with increasing the lower fit boundary and reaches a plateau at $-\ln(\mathrm{JP})_{\mathrm{min}} \approx 3$. The reason for this fit-range dependence is most likely the improper description of the $\ln(\mathrm{JP})$ distributions by MC simulations at medium $\ln(\mathrm{JP})$ as discussed in Sect. 5.4.1: the weighted sum of the fit templates cannot describe the data simultaneously at low and large $\ln(\mathrm{JP})$. The same effect is, however, not visible for fits to the b-tagged data sample.

As a solution for the fit-range dependence, the fit results and the corresponding uncertainties for $f_{\mathrm{b}}^{\mathrm{Tag}}$ and f_{b} are averaged over the results in the *plateau region*. The latter is defined by $0 \leq -\ln(\mathrm{JP})_{\mathrm{min}} \leq 6$ for fits to b-tagged data as well as $2 \leq -\ln(\mathrm{JP})_{\mathrm{min}} \leq 7$ for fits to untagged data with $p_{\mathrm{T,ch\ jet}} \leq 20\,\mathrm{GeV}/c$ and $1 \leq -\ln(\mathrm{JP})_{\mathrm{min}} \leq 6$ for fits to untagged data with $p_{\mathrm{T,ch\ jet}} > 20\,\mathrm{GeV}/c$. These intervals have been chosen based on the observation that the b-jet fractions as well as their statistical uncertainties are about constant in these ranges. Systematic uncertainties that arise from the choice of the plateau regions are discussed in Sect. 5.6.3.

The effect of the fit-range averaging on the quality of the description by the fit model is visible from Fig. 5.12. The latter shows the ratios of the fit model over the data for all seven $p_{\mathrm{T,ch\ jet}}$ bins and lower bin boundaries of $-\ln(\mathrm{JP})_{\mathrm{min}} = 0$ as well as $-\ln(\mathrm{JP})_{\mathrm{min}} = 4$. It can be observed that for low $p_{\mathrm{T,ch\ jet}}$ and $\ln(\mathrm{JP})_{\mathrm{min}} = 0$, the fit model overestimates the untagged data in the interval $8 \lesssim -\ln(\mathrm{JP}) \lesssim 12$ while the interval $16 \lesssim -\ln(\mathrm{JP}) \lesssim 25$ is underestimated. With increasing the lower fit boundary to $-\ln(\mathrm{JP})_{\mathrm{min}} = 4$, the difference between model and data is significantly decreased at medium and large $-\ln(\mathrm{JP})$. This modification becomes negligible for $p_{\mathrm{T,ch\ jet}} > 30\,\mathrm{GeV}/c$ and no modification is apparent for the b-tagged data sample.

The concept of fit-range averaging also has a remarkable effect on the behaviour of the performance-corrected b-jet spectrum and the tagging efficiency, which can be observed in Fig. 5.13. For the performance-corrected spectrum, a significant increase is apparent for $10 < p_{\mathrm{T,ch\ jet}} < 20\,\mathrm{GeV}/c$:
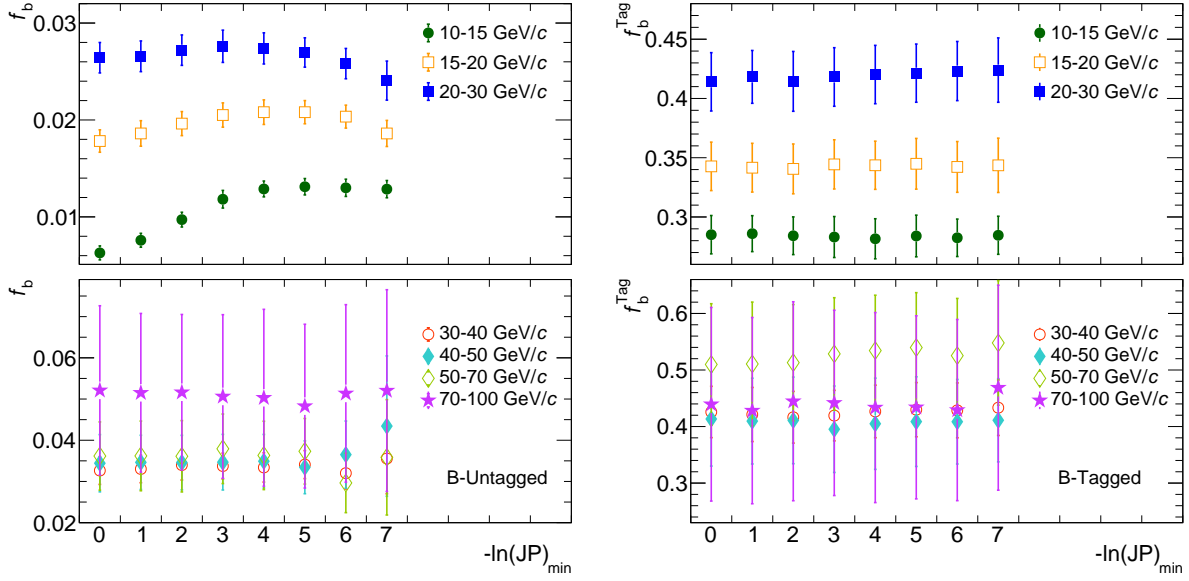
Figure 5.11. – Comparison of the fit results for the b-jet fraction f_b in the untagged data sample (left) and f_b^Tag in the tagged data sample (right) for different $p_\text{T,ch jet}$ bins in dependence of the lower fit boundary $\ln(\text{JP})_\text{min}$.

the b-jet spectrum with the performance obtained from fit-range averaging is larger by up to a factor of two. This modification can be assigned to an increased b-jet fraction in the untagged data sample and a consequential decrease of the tagging efficiency. No such impact is visible for the tagging purity.

Consistency check   To test the accurate implementation of the template fitting procedure as well as the performance correction, a *MC closure test* is performed, which exploits the knowledge of the true b-jet fraction in the MC simulation. The efficiency and purity are obtained from template fits to the MC pseudo data and the raw tagged pseudo data is corrected as defined in Eqn. 5.0.1. The resulting performance-corrected pseudo data is divided by the true detector-level b-jet spectrum, that can directly be obtained from MC. The corresponding ratio can be seen in Fig. 5.14 and is compatible with unity. This proves the correct implementation of the procedure for the performance estimation and concludes the efforts with respect to the quality assurance of the latter.

5.4.4. Results of the performance estimation

Fig. 5.15 compares the results for the performance of the b-jet tagger for PYTHIA 8+GEANT simulations and the ALICE data. The efficiency for the default tagger setting with simultaneous threshold cuts on N1- and N2-type tracks within jets is constant at about 60 % for a purity of 40 % for $20 < p_\text{T,ch jet} < 100\,\text{GeV}/c$. With this, a b-jet tagging efficiency and purity are reached that are of the same order as for the ALICE measurement for pp collisions at $\sqrt{s} = 5.02\,\text{TeV}$ presented in Sect. 4. Both the purity and the efficiency drop for $p_\text{T,ch jet} < 20\,\text{GeV}/c$ due to the low abundance of beauty jets with well-defined $\ln(\text{JP})$ in this region of phase space. The ratio of the tagging purity over the efficiency is about constant for the full $p_\text{T,ch jet}$ range (see Fig. 5.15e). It becomes apparent from Fig. 5.15c and Fig. 5.15d, that the MC simulations predict a larger b-jet fraction in the tagged and the untagged sample for low and medium $p_\text{T,ch jet}$. On the other hand, the simulations underestimate the tagging efficiency in Fig. 5.15b at low $p_\text{T,ch jet}$ by up to almost 30 %. Consequently, the simulations predict a performance-corrected b-jet spectrum shown in Fig. 5.15a which overshoots the data at low $p_\text{T,ch jet}$ by up to a factor of
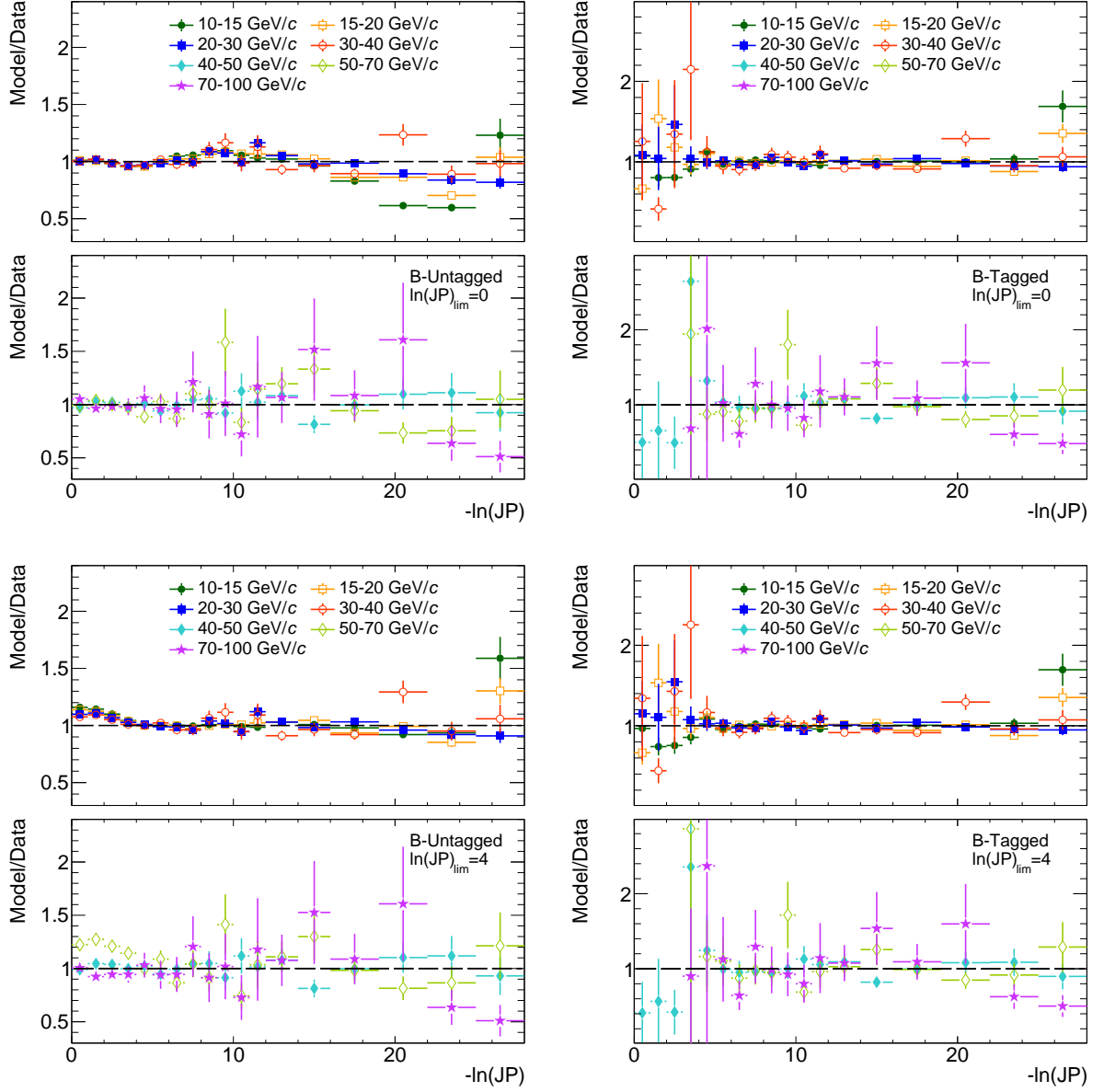
Figure 5.12. – The ratios of the fit models over the data for all seven $p_{\mathrm{T,\,jet}}$ bins of the b-jet spectrum. The results for lower boundaries of $-\ln(\mathrm{JP})_{\mathrm{min}} = 0$ (upper row) and $-\ln(\mathrm{JP})_{\mathrm{min}} = 4$ (lower row) of the fit range are shown.

(a) performance-corrected (pc) spectra

(b) tagging efficiency

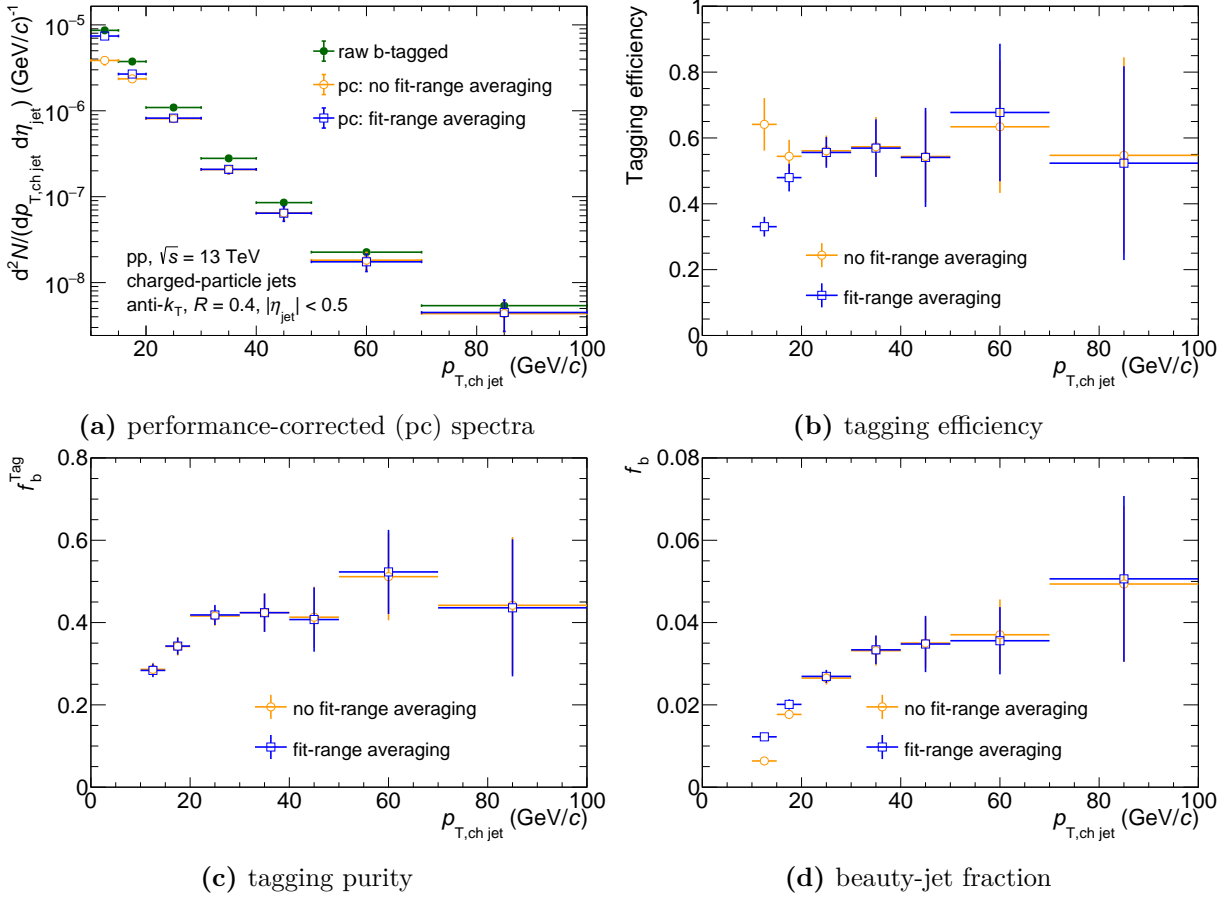**(c)** tagging purity

(d) beauty-jet fraction

Figure 5.13. – Comparison of **a)** the performance-corrected (pc) spectra, **b)** the tagging efficiency ϵ_b, **c)** the tagging purity $f_{\text{b}}^{\text{Tag}}$ and **d)** the b-jet fraction in the untagged sample f_{b} with and without averaging the fit results of different fit ranges. Also the raw b-tagged spectrum is shown for a).
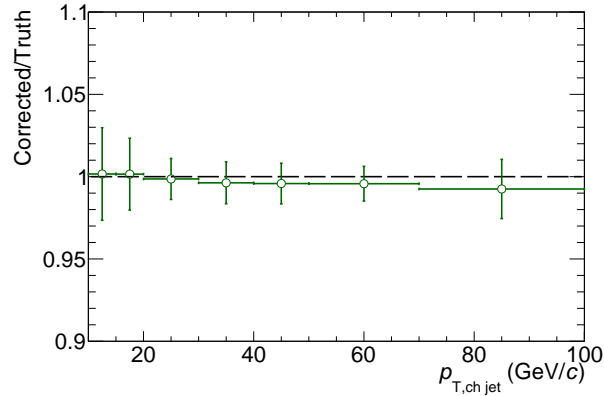

Figure 5.14. – The performance-corrected pseudo data from a MC study divided by the respective detector-level truth (MC Closure test).

about two. It is interesting to compare the behaviour of the performance-corrected spectrum with respect to the raw tagged spectrum for data and MC: Whereas the MC raw pseudo data is performance-corrected upwards for the two lowest $p_{\mathrm{T,ch\ jet}}$ bins, the raw spectrum of the data is corrected downwards. These results indicate that in data, there are more false positive than false negative tags, whereas in MC it is the opposite. This behaviour most probably points towards inaccuracies with respect to the modeling of the positive tails of the Sd_{xy} distributions for the individual jet flavours in the MC production. Several discrepancies are possible in this context: the Sd_{xy} distributions for beauty jets, for example, could reach to larger positive values in data in comparison to the MC simulation. Anyhow, this finding underlines the necessity to determine the purity and efficiency of the b-jet tagger based on a data-driven approach. In addition, it demonstrates that the systematic uncertainties with respect to the MC description of the ln(JP)-template shapes need to be investigated with care, as it is done in Sect. 5.6.1.

5.5. Unfolding of detector effects

To measure the unbiased b-jet cross section, which can be compared to theoretical models and other measurements, the performance-corrected data needs to be corrected for detector inefficiencies with respect to the track and jet reconstruction as well as the momentum reconstruction. The process of eliminating these effects to obtain the *true* physical spectrum is called *unfolding*. The methods to optimise the unfolding procedure that are employed in this thesis correspond to standard procedures within the ALICE Collaboration. In detail, the method of Iterative Bayesian unfolding is applied, which is implemented in the RooUnfold package [28] and will be introduced in the following.

5.5.1. Iterative Bayesian Unfolding

The method of Iterative Bayesian Unfolding [83] is based on Bayes Theorem [59] and allows the calculation of the conditional probability $P_\beta(C_i|E_j)$ that some given effects $E_j \in [1, \ldots, n_E]$ can be assigned to causes $C_i \in [1, \ldots, n_C]$ by

$$P_\beta(C_i|E_j) = \frac{P(E_j|C_i)P_{\beta-1}(C_i)}{\sum\limits_{l=1}^{n_C} P(E_j|C_l)P_{\beta-1}(C_l)}. \tag{5.5.1}$$

Here, $P(E_j|C_i)$ is the probability that the given causes C_i produce effects E_j. In the context of this thesis, the causes C_i can be identified with the true *particle-level* b jets in the QCD vacuum per bin i of their transverse momentum $p_{\mathrm{T,ch\ jet}}^{\mathrm{gen}}$. The effects E_j correspond to the *detector-level* b jets per bin j of their transverse momentum $p_{\mathrm{T,ch\ jet}}^{\mathrm{rec}}$, which have been propagated through the ALICE detector and are thus smeared by detector effects. The probability $P(E_j|C_i)$ is often referred to as the *response matrix* and is obtained, as well as the *prior distribution* $P_0(C_i)$, from MC simulations. The prior distribution can be seen as the best a priori prediction of the true physical spectrum, which is refined to $P_\beta(C_i)$ based on the measured distribution $n(E_j)$ and the response matrix in each iteration β of the unfolding. For every iteration, the predicted number of true causes $n_\beta(C_i)$ and the respective probability density $P_\beta(C_i)$ are calculated as follows

$$n_\beta(C_i) = \frac{1}{\epsilon_i} \sum_{j=1}^{n_E} n(E_j)P_\beta(C_i|E_j) \quad \text{and} \quad N_{\mathrm{true}} = \sum_{i=1}^{n_C} n_\beta(C_i), \tag{5.5.2}$$

$$P_\beta(C_i) = \frac{n_\beta(C_i)}{N_{\mathrm{true}}} \tag{5.5.3}$$

(a) performance-corrected (pc) spectra

(b) tagging efficiency

(c) tagging purity

(d) beauty-jet fraction
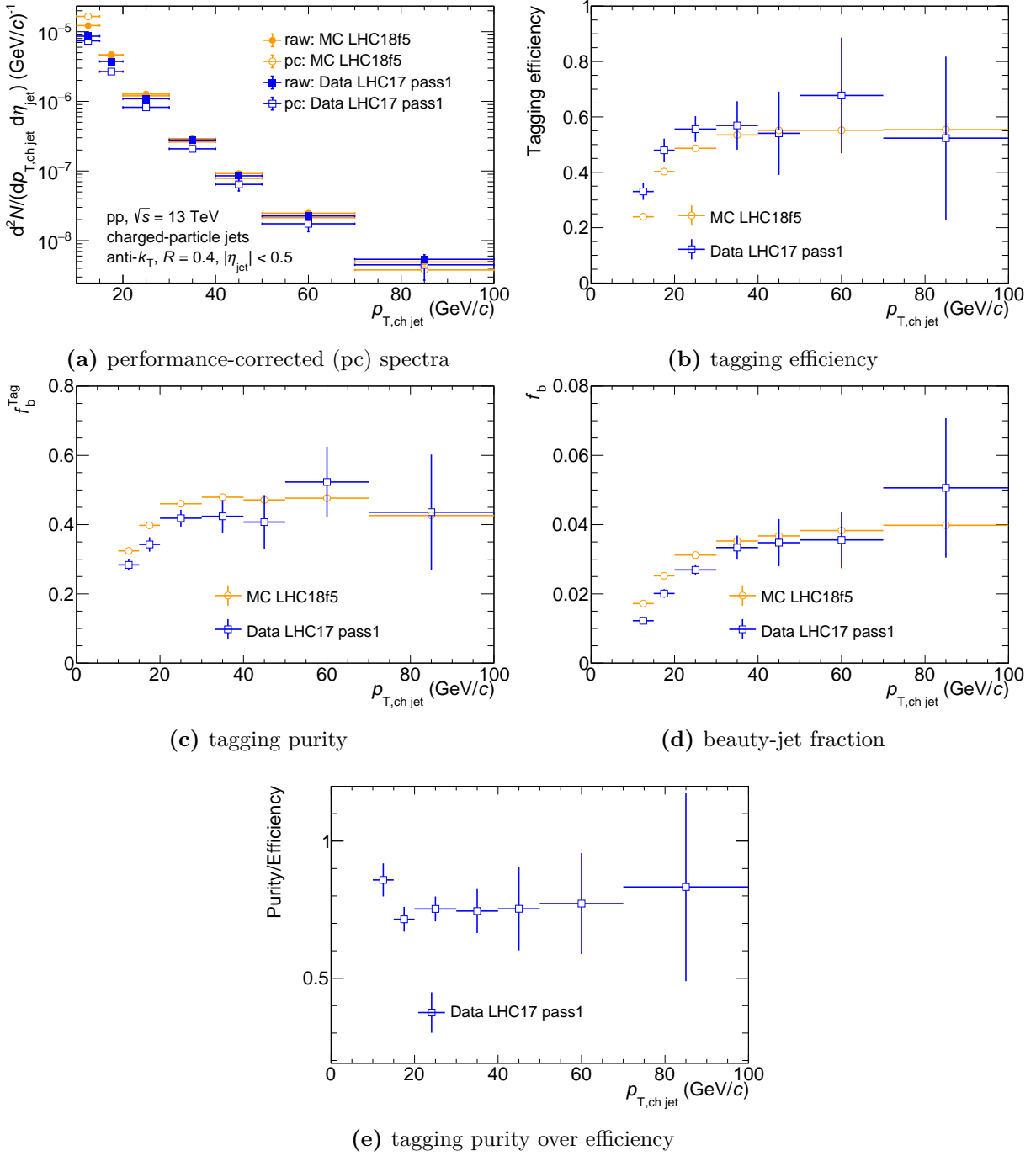
(e) tagging purity over efficiency

Figure 5.15. – Comparison of **a)** the performance-corrected (pc) spectra, **b)** the tagging efficiency ϵ_b, **c)** the tagging purity f_b^{Tag} and **d)** the b-jet fraction in the untagged sample f_b for PYTHIA+GEANT simulations and ALICE data. Also the raw b-tagged spectrum is shown for a). The ratio of the tagging purity over the tagging efficiency for the data is shown in **e)**.
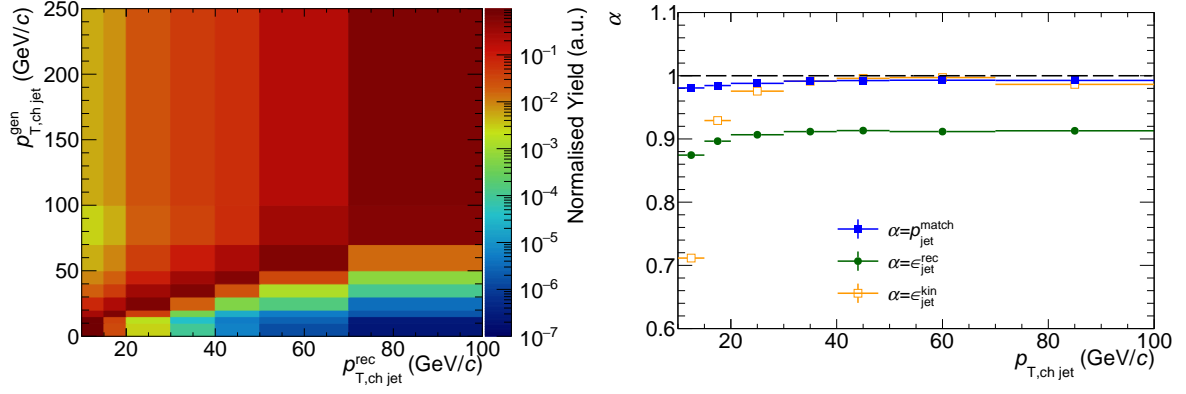
Figure 5.16. – **Left:** The response matrix of the unfolding procedure which provides the number of matched particle-level and detector-level jets per bin of the respective jet momentum for particle-level jets $p_{\mathrm{T,ch\ jet}}^{\mathrm{gen}}$ and for detector-level jets $p_{\mathrm{T,ch\ jet}}^{\mathrm{rec}}$. **Right:** The jet-reconstruction efficiency $\epsilon_{\mathrm{jet}}^{\mathrm{rec}}$, the jet-matching purity $p_{\mathrm{jet}}^{\mathrm{match}}$ and the kinematic efficiency $\epsilon_{\mathrm{jet}}^{\mathrm{kin}}$ in dependence of $p_{\mathrm{T,ch\ jet}}$.

with the unfolding efficiency ϵ_i. The resulting $P_\beta(C_i)$ is then fed to Eqn. (5.5.1) for the next iteration. With increasing number of iterations, the distribution $n_\beta(C_i)$ is assumed to approach the true physical distribution.

5.5.2. Details on the unfolding procedure

The first step in the unfolding procedure is the determination of the response matrix, which relates detector-level beauty jets to their corresponding particle-level counterparts. In this thesis, the response matrix as well as the b-jet prior distribution is obtained from the LHC18f5 MC production.

For a perfect detector, every particle-level jet would produce a respective detector-level jet and the response matrix would be a diagonal matrix. However, due to the detector inefficiencies this assignment suffers from the smearing of the detector-level jets which evokes off-diagonal elements. Particle-level b jets are *matched* to detector-level b jets by geometric criteria i.e. only the pairs that are the closest in angular distance

$$d = \sqrt{(\eta_{\mathrm{det}} - \eta_{\mathrm{part}})^2 + (\varphi_{\mathrm{det}} - \varphi_{\mathrm{part}})^2} \tag{5.5.4}$$

with a maximum value of $d = 0.3$ are filled into the response matrix. Thereby, η_p and φ_p are the pseudorapidity and azimuthal angle of the detector-level ($p = \mathrm{det}$) and the particle-level jet ($p = \mathrm{part}$).

The binning of the response matrix, the performance-corrected data and the unfolded spectrum needs to be adjusted carefully. In particular, the range of the projection of the response matrix on the particle-level truth as well as the range of the prior need to be larger than for the smeared data to account for bin-migration effects. The following binning has been chosen for this analysis:

- smeared data: 10,15,20,30,40,50,70,100 GeV/c,

- particle-level truth: 0,10,15,20,30,40,50,70,100,250 GeV/c.

The response matrix is scaled such that the probability for each particle-level jet to produce a detector-level counterpart is one (i.e. the integral of every row is one). The final response matrix is displayed in Fig. 5.16.

Due to the imperfect matching of particle- and detector-level jets, several corrections need to be applied. If in this context jets are asked to be *within the detector acceptance*, those jets are

required to fulfill $|\eta| < 0.9$ as well as $10 < p_{\mathrm{T,ch\ jet}} < 100\,\mathrm{GeV}/c$. The following corrections have been performed:

Jet-matching purity correction Before the smeared b-jet spectrum is fed to the unfolding procedure, it is corrected for matched detector-level jets which do not have a particle-level counterpart within the detector acceptance. The respective matching purity is defined as

$$p_{\mathrm{jet}}^{\mathrm{match}} = \frac{N_{\mathrm{acc}}^{\mathrm{PL,match}}}{N_{\mathrm{acc}}^{\mathrm{DL,match}}} \tag{5.5.5}$$

with $N_{\mathrm{acc}}^{\mathrm{PL,match}}$ being the number of matched particle-level jets within the detector acceptance and $N_{\mathrm{acc}}^{\mathrm{DL,match}}$ the number of matched detector-level jets with at least one constituent that fulfills the track selection criteria for the b-jet tagging listed in Sect. 5.2.2.

Jet-matching efficiency correction The unfolded b-jet spectrum is corrected by the jet reconstruction efficiency

$$\epsilon_{\mathrm{jet}}^{\mathrm{rec}} = \frac{N_{\mathrm{acc}}^{\mathrm{DL,match}}}{N_{\mathrm{acc}}^{\mathrm{PL}}}, \tag{5.5.6}$$

whereby $N_{\mathrm{acc}}^{\mathrm{PL}}$ is the total number of particle-level jets within the detector acceptance. This efficiency accounts for the number of particle-level jets within the acceptance that get lost during the jet reconstruction. It also eliminates the effect of "fake jets" which are particle-level jets out of the acceptance that are reconstructed within the acceptance.

Kinematic-efficiency correction The unfolded b-jet spectrum is also corrected for the kinematic efficiency of the unfolding

$$\epsilon_{\mathrm{jet}}^{\mathrm{kin}} = \frac{N_{\mathrm{Trunc}}^{\mathrm{PL,match}}}{N_{\mathrm{Untrunc}}^{\mathrm{PL,match}}} \tag{5.5.7}$$

with the number $N_{\mathrm{Untrunc}}^{\mathrm{PL,match}}$ of matched particle-level jets with a detector-level counterpart with $0 < p_{\mathrm{T,ch\ jet}} < 250\,\mathrm{GeV}/c$ and with the number $N_{\mathrm{Trunc}}^{\mathrm{PL,match}}$ of matched particle-level jets with a detector-level counterpart with $10 < p_{\mathrm{T,ch\ jet}} < 100\,\mathrm{GeV}/c$. This efficiency corrects for the fact that particle-level jets with a detector-level jet out of the final $p_{\mathrm{T,ch\ jet}}$-range are not considered by the unfolding procedure.

The smaller the kinematic efficiency, the larger is the dependence of the final results on the MC simulations the response matrix is built on. In Fig. 5.16, one can see that for the lowest $p_{\mathrm{T,ch\ jet}}$ bin, the kinetic efficiency drops to about 70 %, while it is above 90 % for all other bins.

The size of these corrections to the performance-corrected b-jet spectrum $\frac{\mathrm{d}N_{\mathrm{ch\ b\text{-}jet}}^{\mathrm{pc}}}{\mathrm{d}p_{\mathrm{T,ch\ jet}}}$ is shown in Fig. 5.16 and can be combined into one formula

$$\frac{\mathrm{d}N_{\mathrm{ch\ b\text{-}jet}}^{\mathrm{pc}}}{\mathrm{d}p_{\mathrm{T,ch\ jet}}} \rightarrow \mathrm{Unfold}\left(\frac{\mathrm{d}N_{\mathrm{ch\ b\text{-}jet}}^{\mathrm{pc}}}{\mathrm{d}p_{\mathrm{T,ch\ jet}}} p_{\mathrm{jet}}^{\mathrm{match}}\right) \frac{1}{\epsilon_{\mathrm{jet}}^{\mathrm{rec}} \epsilon_{\mathrm{jet}}^{\mathrm{kin}}}. \tag{5.5.8}$$

The statistical uncertainties of the unfolding are determined by varying the bin content of the performance-corrected spectrum within its statistical uncertainties according to a Gaussian distribution. For each of the corresponding toys, the unfolding is performed and the mean and the standard deviation of every $p_{\mathrm{T,ch\ jet}}$ bin are taken as the central value and the uncertainty of the respective bin of the unfolded spectrum.

5.5.3. Convergence and stability tests

The performance of the unfolding procedure can be tested based on MC simulations. Two independent samples of the LHC18f5 MC production are used for this: one for the generation of

the response matrix and one for the testing. Thereby, the statistics of the test sample is chosen to represent the statistics in the real data and the uncertainties are scaled to the uncertainties in the data. The detector-level jets of the test sample are referred to as *pseudo data* and the corresponding particle-level jets as *pseudo truth*.

According to the concept of Iterative Bayesian unfolding, the unfolded pseudo data need to converge to the respective pseudo truth for a sensible number of unfolding iterations β. A basic test (*Closure test I*) can be performed to check whether this objective is met: the ratio of the pseudo truth over the unfolded pseudo data for an increasing number of iterations should approach unity. As it can be seen in Fig. 5.17a, this is the case for the MC test sample from about $\beta = 2$ ongoing.

The stability of the unfolding procedure can be investigated further by reweighting pseudo data and pseudo truth as well as the prior distribution. For every modification the convergence to unity is required. By this, the dependence of the unfolding procedure on the characteristics of the PYTHIA+GEANT simulation can be examined. The following three tests are performed in this respect:

Closure test II: The pseudo data are scaled to match the shape of the distribution of the performance-corrected real data. The ratio of the MC pseudo data over the real data, which has been used for this scaling, can be seen in Fig. 5.17f. The pseudo truth is scaled with the same ratio.

Closure test III: The prior distribution is reweighted to match the shape of the distribution of the original data using the same ratio as it is used in Closure test II. Technically, this entails a corresponding reweighting of the response matrix.

Closure test IV: The prior distribution is rescaled to mimic a POWHEG+PYTHIA 8 simulation. This is done by scaling the default PYTHIA 8 prior by the ratio of simulations for the b-jet distribution in 5.02 TeV pp collisions for POWHEG+PYTHIA 8 with the POWHEG dijet package [40] over the respective results for simulations with PYTHIA 8. The distributions at 5.02 TeV have been provided by the authors of [21] and the respective ratio can be seen in Fig. 5.17e. For reasons that concern the intrinsic settings of POWHEG, the lower boundary of the prior is increased to 3 GeV/c to decrease the influence of threshold behaviour, as described in [40].

As it can be observed in Fig. 5.17, the ratio of the unfolded pseudo data over the pseudo truth for $\beta \geq 3$ is compatible with unity within 10 % deviation for all closure tests. These stability tests based on MC simulations are thus considered to be passed successfully.

Moving to the unfolding of the *real* smeared data, further tests have been performed to assure the reliability of the final results. In particular, the refolded data has to match the original smeared data at every iteration of the unfolding procedure as otherwise the solution is mathematically inconsistent. Since the ratio of the refolded data over the performance-corrected data is compatible with unity within the uncertainties (see Fig. 5.18a), this *refolding test* is positive.

In addition to this, the convergence of the unfolded real data with increasing β needs to be confirmed. In Fig. 5.18b, the ratios of the unfolded spectra for different β over the unfolded spectrum for $\beta = 4$ are presented. This figure demonstrates that the results for the unfolding are converging at $\beta = 4$.

5.5.4. Choosing the regularisation parameter

The optimal number of iterations β that is used for the unfolding procedure evokes the following properties of the corresponding unfolded spectrum:

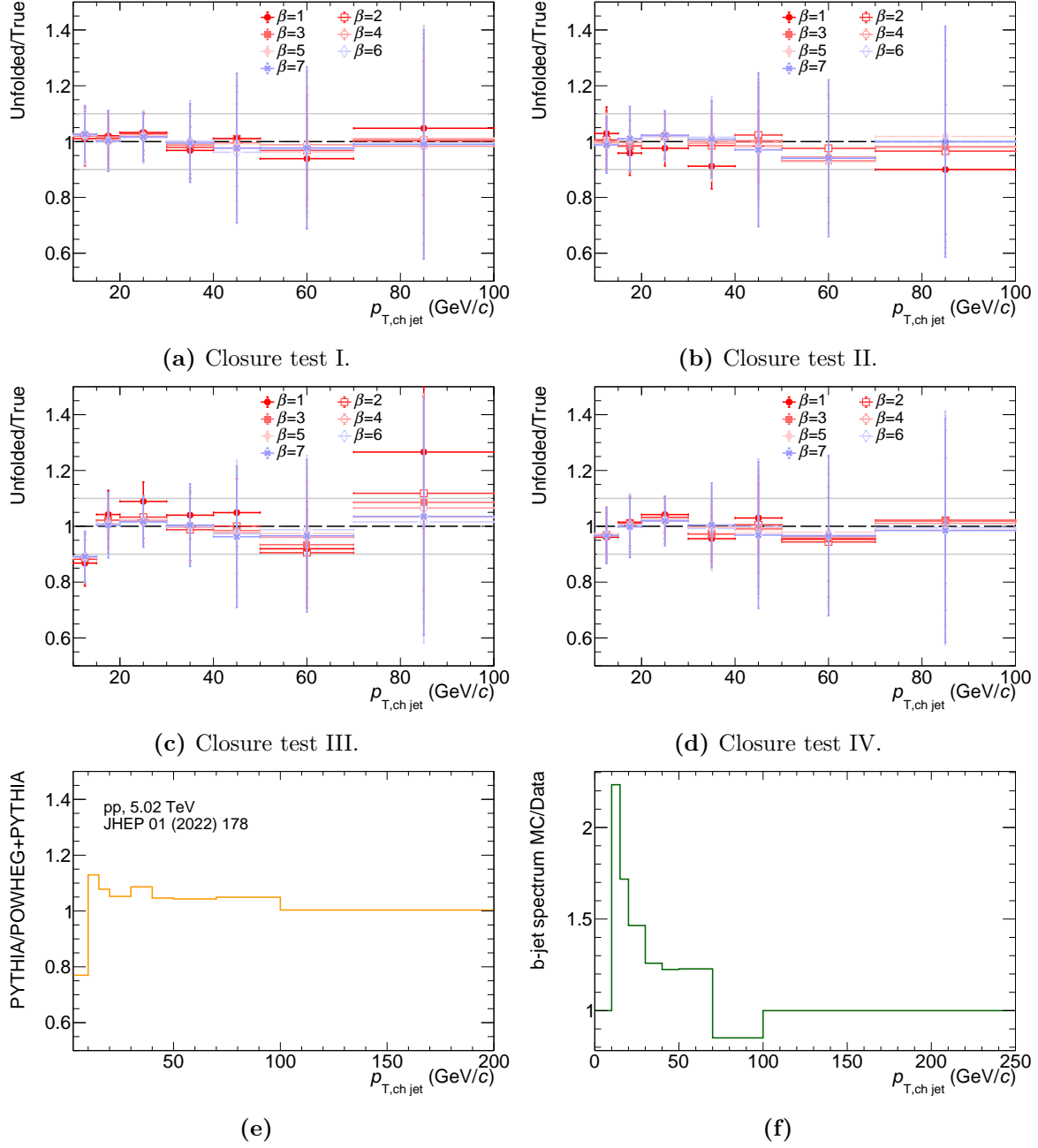1. convergence of the unfolding of the real data within 10 %,

(a) Closure test I.

(b) Closure test II.

(c) Closure test III.

(d) Closure test IV.

(e)

(f)

Figure 5.17. – **a)-d)** The results of the stability tests of the unfolding procedure for different iteration numbers. See text for more detail. **e)** The ratio for POWHEG+PYTHIA dijet over PYTHIA 8 at $\sqrt{s} = 5.02\,\text{GeV}$ that is used for Closure test IV and is taken from [21]. **f)** The ratio of the detector-level b-jet spectrum in MC over the performance-corrected b-jet spectrum in data which is used for Closure Test II.
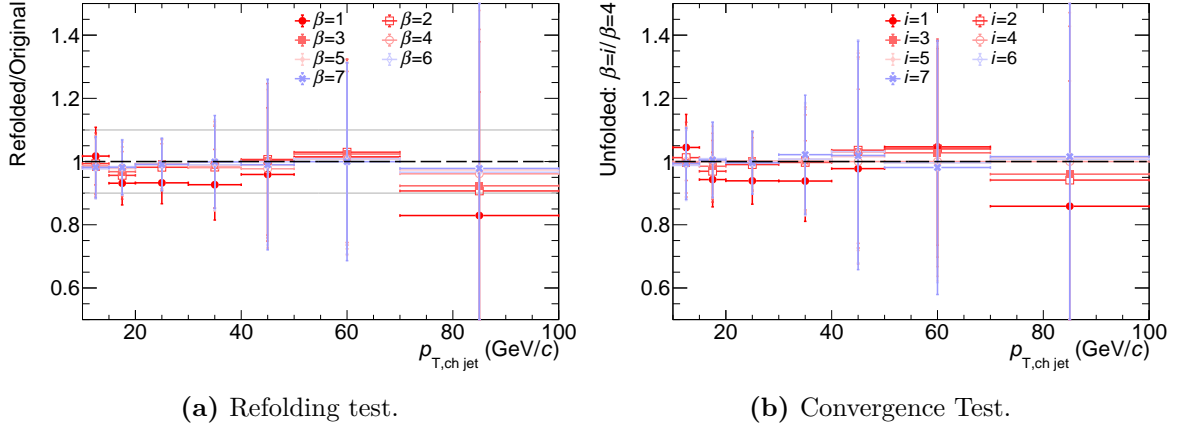
(a) Refolding test. **(b)** Convergence Test.

Figure 5.18. – **a)** The ratio of the refolded data over the original performance-corrected data for different iteration numbers. **b)** The ratio of the unfolded spectra for different iteration numbers over the unfolded spectrum for $\beta = 4$.

2. minimisation of the systematic uncertainties with respect to the choice of the prior distribution and the choice of the iteration number,

3. no visible bin-to-bin correlations of the unfolded spectrum.

Whereas the first two points suggest large iteration numbers, bin-to-bin correlations increase with increasing β. Thus, a reasonable compromise needs to be found.

The results with respect to the first argument have already been presented in the previous section, where it has been found that convergence is given for $\beta \approx 3 - 4$.
To examine the behaviour of the unfolding uncertainties with the number of iterations, the total *prior uncertainty* is defined as a function of β according to the sum over all bins i of the unfolded spectrum

$$\sigma_{\text{prior}}(\beta) = \sum_{i=1}^{n_C} \frac{|n(C_i)^{\text{PY}} - n(C_i)^{\text{POW}}|}{n(C_i)^{\text{PY}}} \qquad (5.5.9)$$

with the unfolded b-jet spectrum $n(C_i)^{\text{PY}}$ that has been obtained using a PYTHIA 8 prior and the unfolded data $n(C_i)^{\text{POW}}$ that has been obtained using a POWHEG+PYTHIA 8 prior.
On the other hand, the total *regularisation uncertainty* in dependence of β is given as the sum over all bins i

$$\sigma_{\text{reg}}(\beta) = \sum_{i=1}^{n_C} \max_{k=+1,-1} \frac{|n(C_i)^{\beta} - n(C_i)^{\beta-k}|}{n(C_i)^{\beta}} \qquad (5.5.10)$$

with the unfolded b-jet spectrum $n(C_i)^{\beta}$ for β iterations.
Fig. 5.19 shows $\sigma_{\text{reg}}(\beta)$ and $\sigma_{\text{prior}}(\beta)$ in comparison to the sum of the statistical uncertainties in dependence of β. The latter uncertainty is significantly larger than $\sigma_{\text{reg}}(\beta)$ and $\sigma_{\text{prior}}(\beta)$. The regularisation uncertainty exhibits a significant decrease at $\beta = 3$ from about 0.3 to 0.1, whereby the prior uncertainty is constant at about 0.1 for all β.

Bin-to-bin correlations are investigated with the help of Pearson matrices, which visualise the correlation coefficient [59]

$$\rho_{ij} = \frac{\text{Cov}(i,j)}{\sigma_i \sigma_j} \qquad (5.5.11)$$

of the bins i and j of the unfolded spectrum. They show now significant off-diagonal correlations for $\beta = 3 - 5$ (see Fig. 5.19).
All above observations are compatible with choosing a number of iterations $\beta = 4$ for the default unfolding procedure.
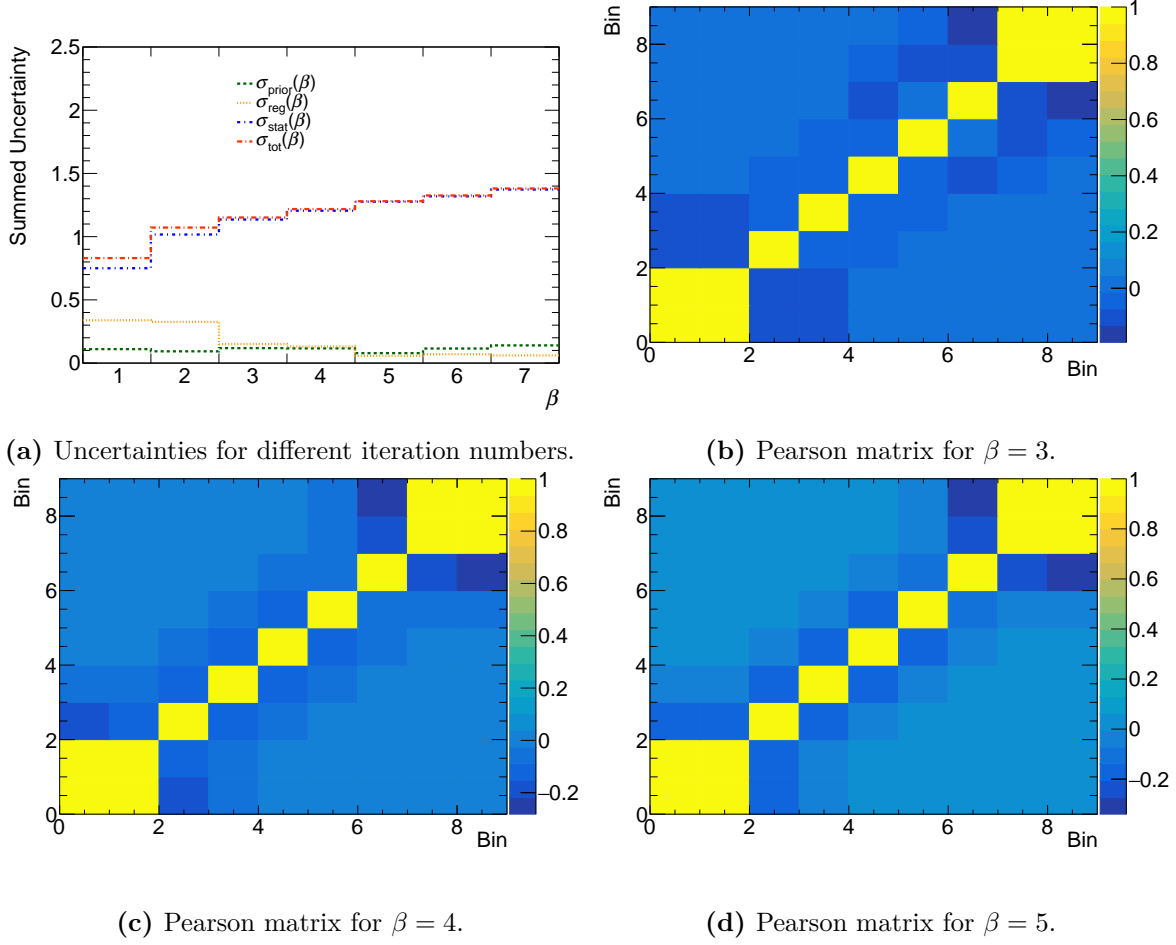
(a) Uncertainties for different iteration numbers.

(b) Pearson matrix for $\beta = 3$.

(c) Pearson matrix for $\beta = 4$.

(d) Pearson matrix for $\beta = 5$.

Figure 5.19. – **a)** The total statistical $\sigma_{\mathrm{stat}}(\beta)$, regularisation $\sigma_{\mathrm{reg}}(\beta)$ and prior uncertainty $\sigma_{\mathrm{prior}}(\beta)$ in dependence of the number of iterations β. The total unfolding uncertainty $\sigma_{\mathrm{tot}}(\beta)$ is given by the quadrature sum of the individual uncertainties. **b)-d)** The Pearson matrices for $\beta = 3, 4$ and 5.

5.6. Estimation of systematic uncertainties

The systematic uncertainties that are considered in this analysis can be split into two groups. The first group is formed by the uncertainties Δ_{unfold} of the unfolding procedure. Uncertainties that arise in the analysis steps before the unfolding procedure will be referred to as *uncertainties on the performance-corrected spectrum* Δ_{pc} in the following. These uncertainties have to be propagated through the unfolding procedure. However, among this group of uncertainties, there are some that need to be unfolded with a dedicated response matrix. These will be called uncertainties of *category 2*, while *category 1* uncertainties are unfolded with the default response matrix.

For every uncertainty that is investigated for the performance-corrected spectrum, the Barlow test [53] is performed. This means that the following observable

$$f_{\mathrm{Bar}} = \frac{x_{\mathrm{Default}} - x_{\mathrm{Sys}}}{\sqrt{|\sigma_{\mathrm{Default}}^2 - \sigma_{\mathrm{Sys}}^2|}} \tag{5.6.1}$$

is calculated for every $p_{\mathrm{T,ch\ jet}}$ bin of the performance-corrected b-jet spectrum. Thereby, x_{Default} corresponds to the bin content of the default spectrum and x_{Sys} to the bin content of the spectrum that is obtained from the systematic variation. $\sigma_{\mathrm{Default}}$ and σ_{Sys} are the corresponding statistical uncertainties. If the test is passed, i.e.

$$|f_{\mathrm{Bar}}| \lesssim 1 \tag{5.6.2}$$

is valid for all bins, the deviation from the default spectrum is classified as statistically insignificant and the respective uncertainty is not included in the calculation of the final systematic uncertainty. Since the systematic variations with respect to the unfolding procedure are well established in the ALICE collaboration, all corresponding variations are considered for the final systematic uncertainty and Barlow tests have not been performed.

The final systematic uncertainty Δ_{sys} is given by the squared sum of the systematic uncertainties from the unfolding procedure Δ_{unfold} and the uncertainties $\Delta_{\mathrm{pc,\ cat1.}}$ and $\Delta_{\mathrm{pc,\ cat.2}}$ on the performance-corrected spectrum from category 1 and 2

$$\Delta_{\mathrm{sys}} = \sqrt{\Delta_{\mathrm{unfold}}^2 + \Delta_{\mathrm{pc,\ cat1.}}^2 + \Delta_{\mathrm{pc,\ cat.2}}^2}. \tag{5.6.3}$$

In the following sections, the implementations of all systematic variations are described. The Barlow tests and the resulting systematic uncertainties from all sources are summarised in Sect. 5.6.5.

5.6.1. Shape uncertainties

In Sect. 5.3.2, it has been shown that the Sd_{xy} distributions which are simulated with PYTHIA and GEANT are in general more narrow than respective distributions in data. The definition of the jet probability $\ln(\mathrm{JP})$ is based on information about the track impact parameter resolution (see Sect. 5.4.1). Thus, discrepancies with respect to the Sd_{xy} distributions in data and MC simulations are propagated to the corresponding $\ln(\mathrm{JP})$ distributions.

In this section, the sensibility of the performance-corrected b-jet spectrum with respect to the shape of the $\ln(\mathrm{JP})$ templates and the data that are used for the performance estimation will be investigated. As described in the following, all shape modifications are based on manipulations of the underlying impact-parameter distributions that are used for the estimation of the $\ln(\mathrm{JP})$.

Variation of the resolution function

The default method for the performance estimation utilises the resolution function $R_{\text{MC}}(I)$ obtained from the LHC18f5 MC production for the ln(JP) calculation in the data and for the MC fit templates. In total, the impact of three systematic variations of the template shapes have been analysed:

- Variation 1 (Var.1): The resolution function $R_{\text{data}}(I)$ from data is used for the ln(JP) calculation both in data and in MC.

- Variation 2 (Var.2): $R_{\text{data}}(I)$ is used for the ln(JP) calculation in data and $R_{\text{MC}}(I)$ for the ln(JP) calculation in MC.

- Variation 3 (Var.3): $R_{\text{data}}(I)$ is used for the ln(JP) calculation in data and in MC. In addition, the impact-parameter distribution $(Sd_{xy})^{\text{N1}}_{fl}$ for N1-type tracks within jets of a particular flavour $fl \in [\text{uds,c,b,incl}]$ in MC is scaled according to the ratio $r_{\text{Data/MC}}$ of the distributions for inclusive jets in data and MC, i.e.

$$(Sd_{xy})^{\text{N1}}_{fl} \rightarrow (Sd_{xy})^{\text{N1}}_{fl} \underbrace{\frac{(Sd_{xy})^{\text{N1}}_{\text{incl,data}}}{(Sd_{xy})^{\text{N1}}_{\text{incl,MC}}}}_{r_{\text{Data/MC}}} . \tag{5.6.4}$$

The respective ratio $r_{\text{Data/MC}}$ is obtained from the ratio of fits with the fit function defined in Sect. A.1 for N1-type tracks in data and MC.

The fit templates that have been modified according to these variations are fed to the template-fitting procedure and the corresponding results for the tagging performance are used to obtain the modified performance-corrected b-jet spectrum.

The fit templates for Var.1 and Var.2 can be obtained using the LHC18f5 MC production. Fig. 5.20 compares the respective ln(JP) distributions for Var.1 to the default distributions for b jets with $10 < p_{\text{T,ch jet}} < 15\,\text{GeV}/c$ and for b-tagged and untagged MC samples. It becomes apparent that the ln(JP) distributions for Var.1 are steeper with respect to the default distributions. At large ln(JP), this leads to a difference of up to 50% in case of the untagged MC sample.

The ln(JP) distributions can also be subject to biases that are caused by the *intrinsic* impact-parameter resolution $R^{\text{in}}_{\text{data}}(I)$ in the data being different from the resolution function $R_{\text{MC}}(I)$ used for the ln(JP) calculation (case 2) in Sect. 5.4.1). This effect can not be studied using available MC productions since the intrinsic impact-parameter resolution would need retuning. Thus, an ln(JP) sampling algorithm has been developed to implement Var.3. This algorithm is described in the next subsection.

ln(JP) sampling algorithm

The ln(JP) sampling algorithm is based on hit-and-miss as well as importance sampling [59, 133] of the probability distributions for jet and track observables. Both sampling techniques constitute numerical methods to generate random numbers following a predefined distribution and are detailed in Appendix C.1. Using these techniques, jet objects are simulated as entities of tracks with respective impact-parameter values. Thereby, the number of tracks and the track impact-parameters are obtained in a way that their distributions follow corresponding probability distributions from the LHC18f5 MC production. The ln(JP) of these jet objects is calculated as before. The final result of the sampling algorithm is an ln(JP) distribution for each specific $p_{\text{T,ch jet}}$ bin which can easily be manipulated based on modifications of the underlying probability distributions.

The following probability distributions serve as input histograms to the ln(JP) sampling algorithm for the simulation of jets with flavour $fl \in [\text{uds,c,b}]$:
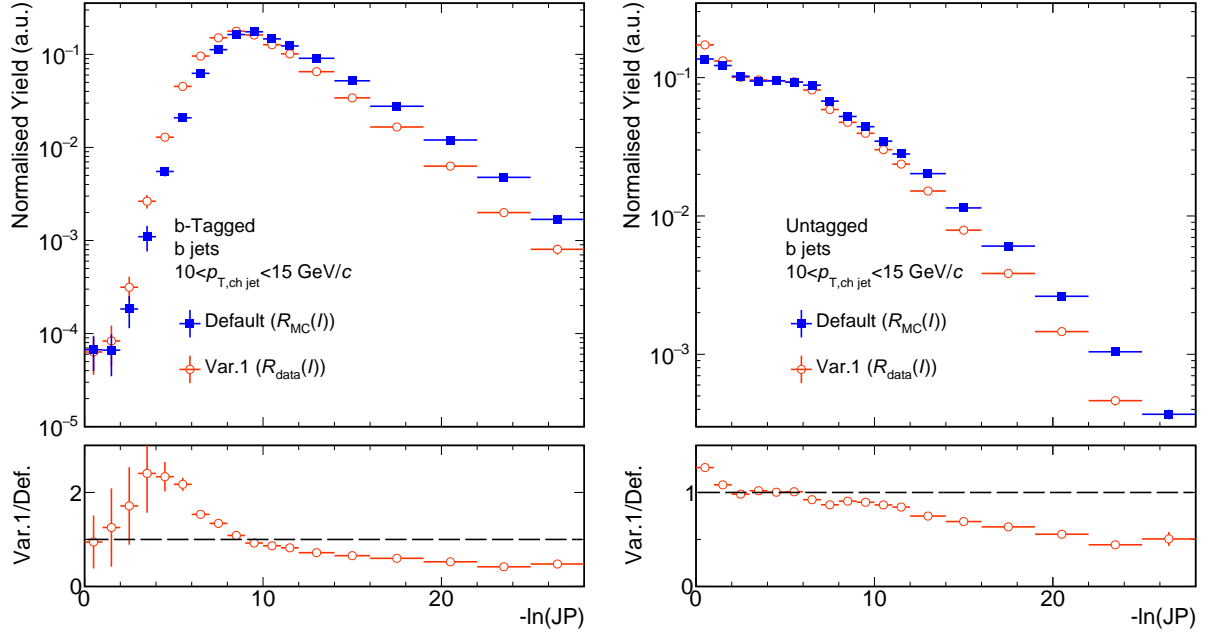
Figure 5.20. – Comparison of the b-jet ln(JP) distributions from the LHC18f5 MC production obtained by the default method for the performance correction with respect the distributions obtained from Var.1. The distributions for b-tagged and b-untagged jets with $30 < p_{\mathrm{T,ch\ jet}} < 40\,\mathrm{GeV}/c$ are shown.
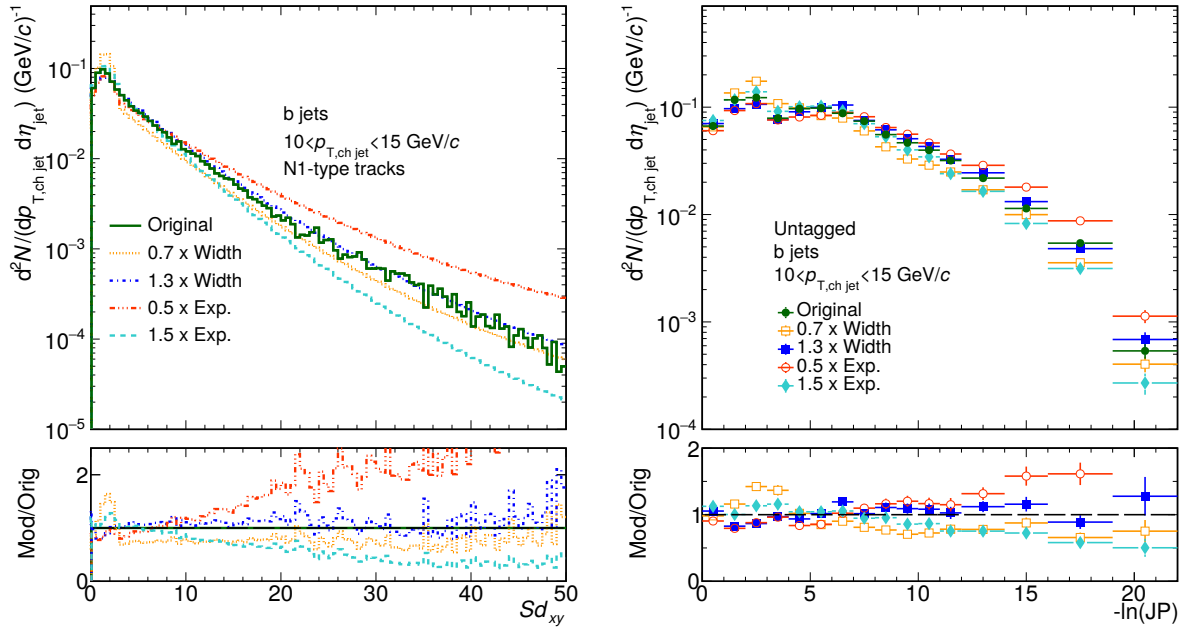

Figure 5.21. – Results for testing the influence of a modification of the intrinsic impact-parameter resolution on the corresponding ln(JP) distributions using the ln(JP) sampling algorithm. **Left:** Respective modifications of the $(Sd_{xy})_{\mathrm{b}}^{\mathrm{N1}}$ distributions for the impact parameter of N1-type tracks within b jets. The modifications correspond to a scaling of the parameters g_3 ("Width") and $t_{\beta,4}$ ("Exp.") of the fit function in Sect. A.1. The corresponding scale factors are provided in the legend. **Right:** Effects on the ln(JP) distributions of the modifications of the $(Sd_{xy})_{\mathrm{b}}^{\mathrm{N1}}$ distributions shown in a).

- the distributions of the impact parameter significance $(Sd_{xy})_{fl}^{\mathrm{N1}}$ for N1-type tracks within jets of flavour fl,

- the two-dimensional correlations of the impact parameters of Ni- and $N(i+1)$-type tracks with $i \leq 4$ for jets of flavour fl,

- the distributions of the number of tracks N_{tr}^{fl} within jets of flavour fl,

- the resolution function $R(I)$ for inclusive jets (taken from data or the LHC18f5 MC production depending on the kind of systematic variation).

The procedure that is used for the generation of a single jet object of flavour fl can be summarised as follows:

1. The number of tracks N_{tr} is obtained from importance sampling off the N_{tr}^{fl} distribution. It determines how many tracks will be generated for a specific jet object.

2. The impact parameter Sd_{xy}^{N1} of the jet's N1-type track is obtained from importance sampling off the $(Sd_{xy})_{fl}^{\mathrm{N1}}$ distribution.

3. The resulting Sd_{xy}^{N1} is used as input for hit-and-miss sampling of the Sd_{xy}^{N2} of the $N2$-type track via the respective correlation histogram. The value Sd_{xy}^{N2} is itself utilised as input for the generation of Sd_{xy}^{N3} and this queue-like sampling is only stopped if the maximum number of tracks is reached. If a jet object exhibits more than five tracks, the correlation histogram for N4- and N5-type tracks are applied.

4. Similar as for the default analysis, a tagging decision is deduced from the generated Sd_{xy} values according to the tagging criteria described in Sect. 5.3.

5. Based on the resolution function and the sampled tuple of impact parameters, the $\ln(\mathrm{JP})$ of the respective jet object can be calculated using Eqn. 5.4.2.

The impact of a modification of $R_{\mathrm{data}}^{\mathrm{in}}(I)$ on the $\ln(\mathrm{JP})$ distributions can be tested using the sampling algorithm. Thereby, $R_{\mathrm{data}}^{\mathrm{in}}(I)$ is represented by the $(Sd_{xy})_{fl}^{\mathrm{N1}}$ distributions and, by modifying the latter, the effect on the $\ln(\mathrm{JP})$ distributions can be investigated. In this sense, the $(Sd_{xy})_{\mathrm{b}}^{\mathrm{N1}}$ distributions for b jets have been fit with the fit function from Sect. A.1 for $(Sd_{xy})_{\mathrm{b}}^{\mathrm{N1}} > 0$. The width of the Gaussian function for the central peak and the level of the tails have been modified and these modified distributions have been fed to the sampling algorithm. The modified $(Sd_{xy})_{\mathrm{b}}^{\mathrm{N1}}$ distributions are shown for $10 < p_{\mathrm{T,ch\ jet}} < 15\,\mathrm{GeV}/c$ together with the respective $\ln(\mathrm{JP})$ distributions in Fig. 5.21. It can be seen that modifications of the Sd_{xy} distributions are directly transfered to the corresponding $\ln(\mathrm{JP})$ distributions although they are smeared over a comparably wide $\ln(\mathrm{JP})$ range. To conclude, the assumption is justified that biases arise if the intrinsic Sd_{xy} resolution of a data sample and the resolution function that is used for the $\ln(\mathrm{JP})$ calculation are different.

The performance of the $\ln(\mathrm{JP})$ sampling algorithm can be tested by comparing its predictions to respective $\ln(\mathrm{JP})$ distributions obtained from the LHC18f5 MC production. In this specific case, the respective $\ln(\mathrm{JP})$ distributions are compared for the default $\ln(\mathrm{JP})$ distributions and for Var.1. The results for light-flavour, c and b jets can be seen in Fig. 5.22. To begin with, the $\ln(\mathrm{JP})$ distributions of the sampling algorithm largely underestimated the $\ln(\mathrm{JP})$ distributions of the LHC18f5 MC production for $\ln(\mathrm{JP}) > 15$ for all jet flavours. However, the different trend of the default distributions compared to those for Var.1 is well described. This indicates, that the $\ln(\mathrm{JP})$ sampling algorithm provides distributions that merely exhibit an offset with respect to the true $\ln(\mathrm{JP})$ distributions. The latter can be eliminated by considering ratios of the $\ln(\mathrm{JP})$ distributions from the sampling algorithm. In this sense, Var.3 is implemented as the scaling of
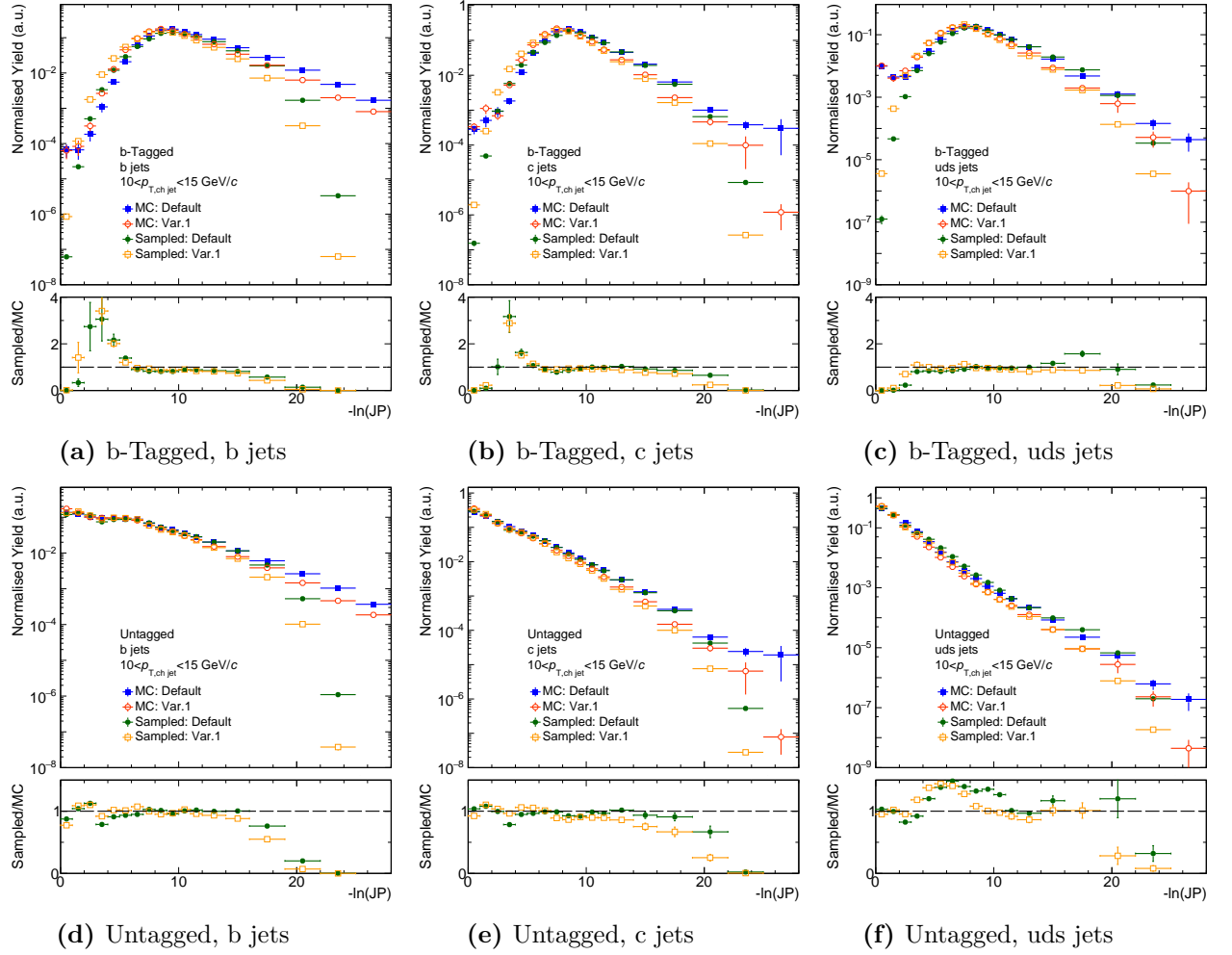
(a) b-Tagged, b jets **(b)** b-Tagged, c jets **(c)** b-Tagged, uds jets

(d) Untagged, b jets **(e)** Untagged, c jets **(f)** Untagged, uds jets

Figure 5.22. – The ln(JP) distributions for b, c and light-flavour jets obtained via the ln(JP) sampling algorithm compared to the distributions from the LHC18f5 MC production. The respective distributions are shown for the default settings and for Var.1. The distributions for the b-tagged data sample are shown in **a)-c)** and those for the untagged data sample in **d)-f)**.
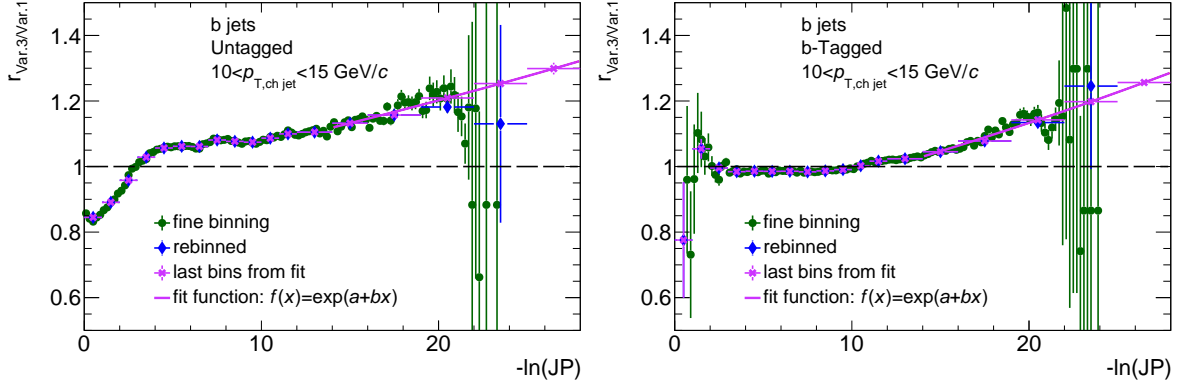
Figure 5.23. – The ratio $r_{\mathrm{Var.3/Var.1}}$ of the ln(JP) distributions for the systematic variations Var.1 and Var.3 for b jets with $10 < p_{\mathrm{T,ch\ jet}} < 15\,\mathrm{GeV}/c$. The distributions are shown for untagged (left) and b-tagged (right) jets together with corresponding fit functions $f(x)$. The ratios that are obtained from calculating the bins contents for ln(JP) < 19 based on $f(x)$ are also provided.

the fit templates in Eqn. 5.4.3 according to

$$
\mathcal{P}_{fl}^{i}(\ln(\mathrm{JP})) \to \mathcal{P}_{fl}^{i}(\ln(\mathrm{JP})) \cdot \underbrace{\frac{\mathcal{P}_{fl}^{i,\mathrm{Var.3}}(\ln(\mathrm{JP}))}{\mathcal{P}_{fl}^{i,\mathrm{Var.1}}(\ln(\mathrm{JP}))}}_{r_{\mathrm{Var.3/Var.1}}} . \tag{5.6.5}
$$

for every bin i with the respective ln(JP) distributions from Var.3 referred to as $\mathcal{P}_{fl}^{i,\mathrm{Var.3}}(\ln(\mathrm{JP}))$ and the distributions for Var.1 referred to as $\mathcal{P}_{fl}^{i,\mathrm{Var.1}}(\ln(\mathrm{JP}))$. The corresponding ratios for b jets in the b-tagged and the untagged MC sample are shown for $10 < p_{\mathrm{T,ch\ jet}} < 15\,\mathrm{GeV}/c$ in Fig. 5.23. They exhibit an increasing trend with increasing ln(JP), which corresponds to ln(JP) distributions for Var.3 which are flatter than those for Var.1. Further examples for the ratios of all jet flavours are provided in Appendix C.1. They show that the ratios approach unity with increasing $p_{\mathrm{T,ch\ jet}}$.

As it can be recognised in Fig. 5.23, the ratios $r_{\mathrm{Var.3/Var.1}}$ suffer from low statistics at large ln(JP) > 20. Therefore, they are fit with an exponential function $f(x) = \exp(a + bx)$ and the bin contents for $-\ln(\mathrm{JP}) > 19$ are calculated from the fit function for the further analysis.

Effects of variations on the template-fitting stability   As for the default process of the performance estimation, template fits with the modified fit templates have been performed to the ln(JP) distributions of the modified data. The impact of the systematic variations with respect to the fit stability can be assessed by looking at the ratio of the fit model over the data for fits to the default ln(JP) range *without* averaging over the results from data toy studies, as described in Sect. 5.4.3. It can also be studied by looking at the b-jet fractions in dependence of the lower fit boundary for fits *with* averaging over the results from data toy studies. The respective distributions for Var.1 and Var.3 are shown in Fig. 5.24 and Fig. 5.25, which can be compared to Fig. 5.12 and Fig. 5.11 for the results of the default method. For Var.1, no significant difference can be observed with respect to the model description and the fit-range dependence of the b-jet fraction in comparison to the default settings. For Var.3, a slight tendency towards a better description of the untagged data for about $10 < \ln(\mathrm{JP}) < 24$ is visible for $p_{\mathrm{T,ch\ jet}} < 30\,\mathrm{GeV}/c$. Also, the fit-range dependence of f_{b} is decreased for $p_{\mathrm{T,ch\ jet}} < 20\,\mathrm{GeV}/c$. These observations point towards a mismatch of the resolution function in data and MC, which is responsible for the instabilities of the template fitting procedure at low $p_{\mathrm{T,ch\ jet}}$ with respect to different fit ranges that was observed in Sect. 5.4.3.

For Var.2, no stable results could be obtained using the template fitting procedure in this

analysis. The most probable reason for this observation is the strong difference of the ln(JP) definitions based on whether the resolution functions $R_{\mathrm{data}}(I)$ or $R_{\mathrm{MC}}(I)$ are applied for the ln(JP) estimation.

Effects on performance observables and final spectrum

Fig. 5.26 shows a comparison of the $p_{\mathrm{T,ch\ jet}}$-dependent b-jet fractions in the untagged and b-tagged data sample for the default method of the performance estimation with respect to the results from Var.1 and Var.3. The results for f_{b} are compatible for all methods, which is remarkable considering the different behaviour with respect to the fit stability of Var.3 discussed in the previous paragraph. This observation, however, indicates that the fit-range averaging, which has been introduced in Sect. 5.4.3, is able to compensate effects from biased template shapes for the untagged jets.

On the other hand, Var.3 predicts a significantly smaller $f_{\mathrm{b}}^{\mathrm{Tag}}$ for $p_{\mathrm{T,ch\ jet}} < 20\,\mathrm{GeV}/c$. Consequently, Var.3 favours a slightly flatter shape for the performance-corrected b-jet spectrum although the final spectra are compatible with each other within the uncertainties (see Fig. 5.27a).

As it was shown in Sect. 5.3.2, the statistics for the Sd_{xy} distributions for $N1$-type tracks is low for data at large $p_{\mathrm{T,ch\ jet}}$. This is why the fits to obtain $r_{\mathrm{data/MC}}$ for the scaling with respect to Var.3 do not provide sensible results for $p_{\mathrm{T,ch\ jet}} > 50\,\mathrm{GeV}/c$. Consequently, the systematic uncertainty with respect to modified resolution functions cannot directly be obtained for larger $p_{\mathrm{T,ch\ jet}}$ and the final results are extrapolated from low to larger $p_{\mathrm{T,ch\ jet}}$. This is done by fitting the uncertainties for $20 < p_{\mathrm{T,ch\ jet}} < 50\,\mathrm{GeV}/c$ with a constant fit function as shown in Fig. 5.27b and replacing the bin content for $50 < p_{\mathrm{T,ch\ jet}} < 100\,\mathrm{GeV}/c$ by the normalised integrals of the fit function.

The performance-corrected b-jet spectrum for Var.3 shows a larger deviation from the default spectrum than Var.1 at low $p_{\mathrm{T,ch\ jet}}$ where the systematic variations are most significant. This is why, Var.3 is considered as the final shape uncertainty for the estimation of the total systematic uncertainties.

A note on prospects for future analysis The procedure that is presented in this section for the estimation of the template-shape uncertainties demonstrates the substantial dependence of the b-tagging performance estimation by ln(JP) template fits on a suitable MC description of the impact-parameter resolution. Although the method is well suited to illustrate this dependency, it also suffers from major drawbacks which concern the very basic assumptions that it is built on. This, for instance, leads to the significant underestimation of the ln(JP) distributions for large ln(JP) values, which could not be solved conclusively in the course of this thesis. Consequently, dedicated detector-level MC simulations for assuming different impact-parameter resolutions would be favourable in investigating the template-shape uncertainties for future analysis. Also, the examination of the fit templates with different particle-level MC generators would provide valuable information for constraining the dependence of the method for the performance estimation on the characteristics of MC event generators.

Apart from this, the method presented here will naturally profit from an extension of the analysis to the full 13 TeV data set of ALICE. With further statistics, the determination of the ratio $r_{\mathrm{Data/MC}}$ might also be possible for further $p_{\mathrm{T,ch\ jet}}$ bins with $p_{\mathrm{T,ch\ jet}} > 50\,\mathrm{GeV}/c$. In this case, the uncertainties of the respective bins could be determined independently without the need for extrapolation from low $p_{\mathrm{T,ch\ jet}}$.

5.6.2. V^0-jet uncertainty

The default method for the performance estimation relies on a sensible MC description of all background candidates. However, PYTHIA 8 shows significant deviations with respect to
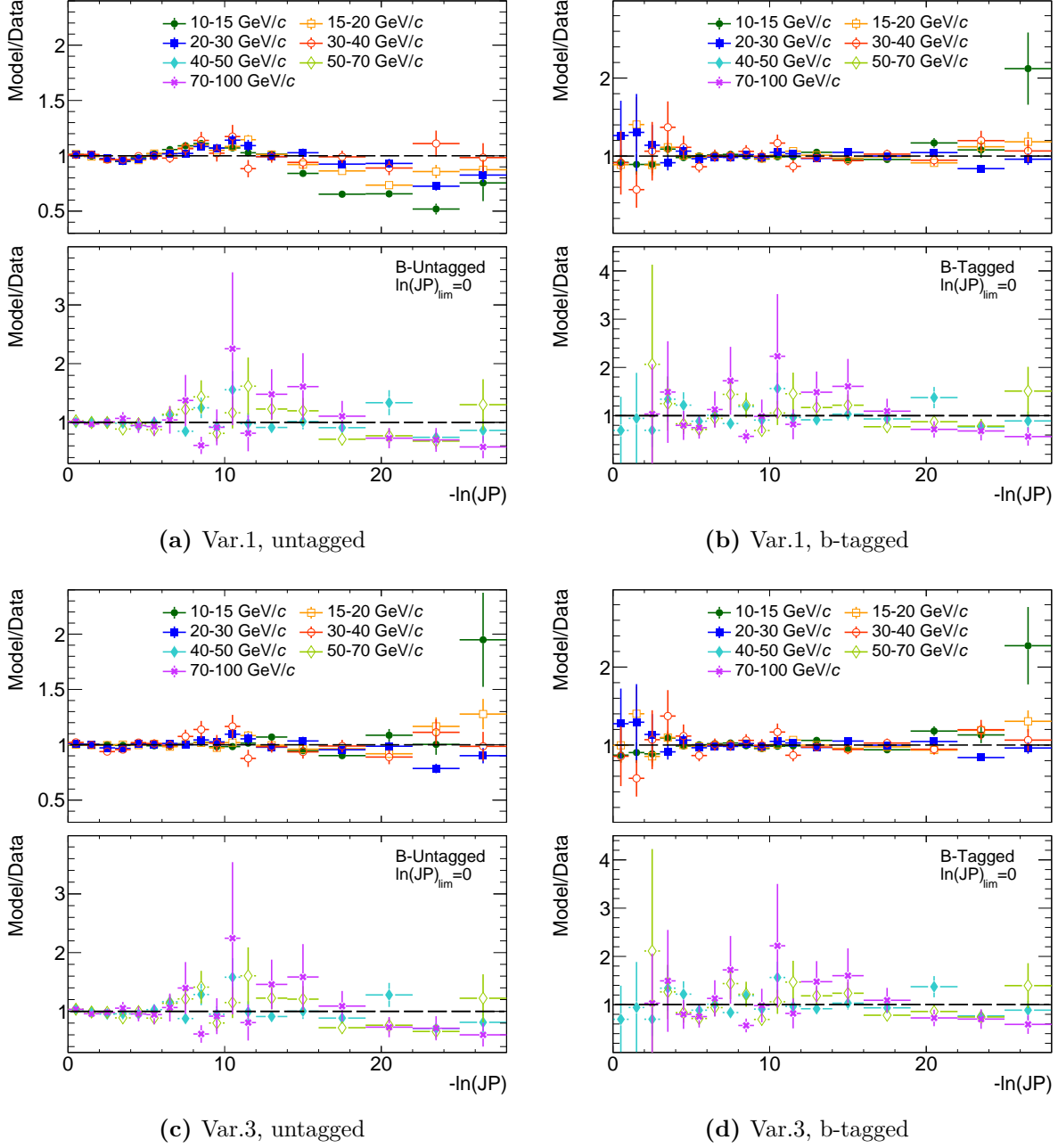
(a) Var.1, untagged

(b) Var.1, b-tagged

(c) Var.3, untagged

(d) Var.3, b-tagged

Figure 5.24. – Ratios of the fit model over the data for the template-fitting procedure with the data and the fit templates modified according to Var.1 ((a) and (b)) and Var.3 ((c) and (d)). See Fig. 5.12 for corresponding distribution of the default procedure.
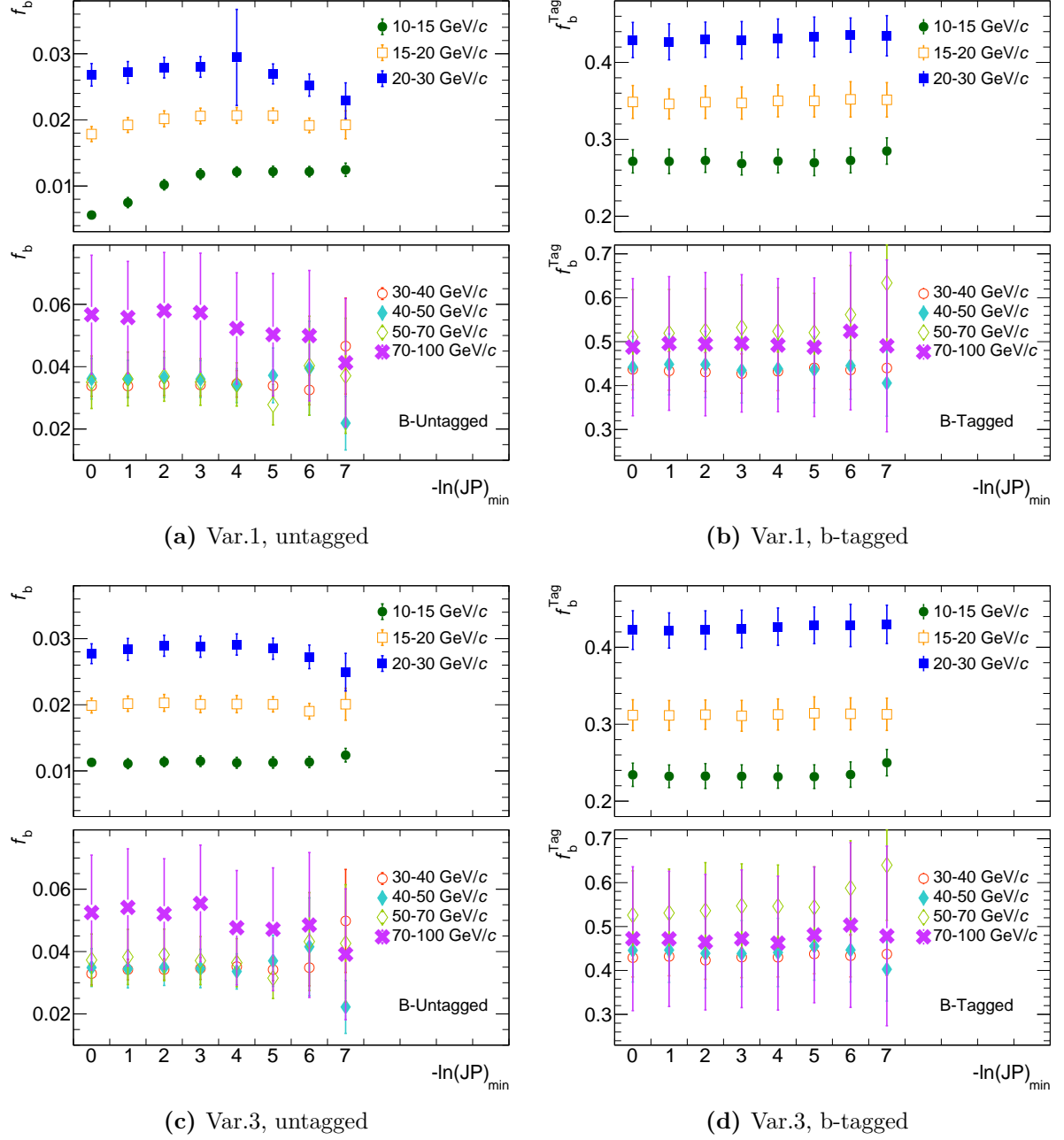
Figure 5.25. – b-jet fractions obtained from the template-fitting procedure for the b-tagged and untagged data samples in dependence of the lower fit boundary for Var.1 ((a) and (b)) and Var.3 ((c) and (d)). See Fig. 5.11 for corresponding distribution of the default procedure.
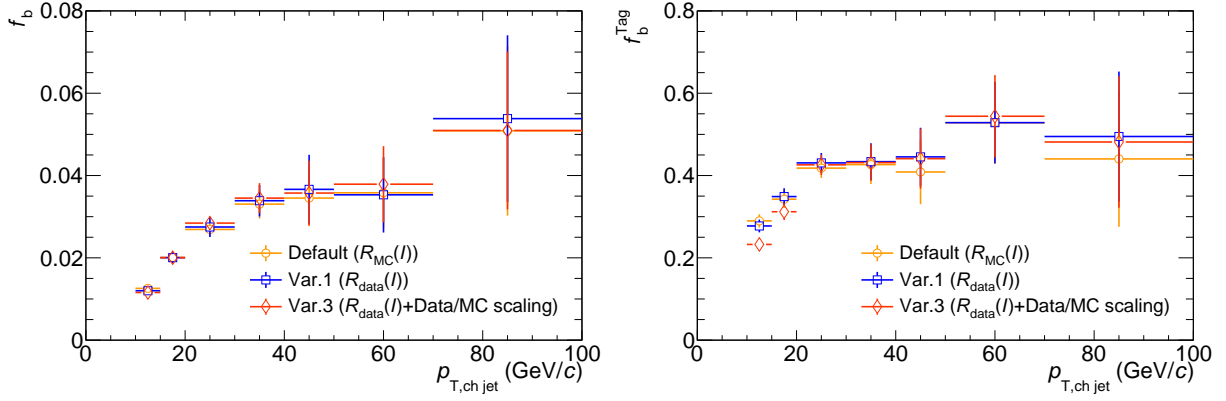
Figure 5.26. – Comparison of the b-jet fractions in the untagged (left) and b-tagged (right) data sample for the default method of performance correction as well as Var.1 and Var.3
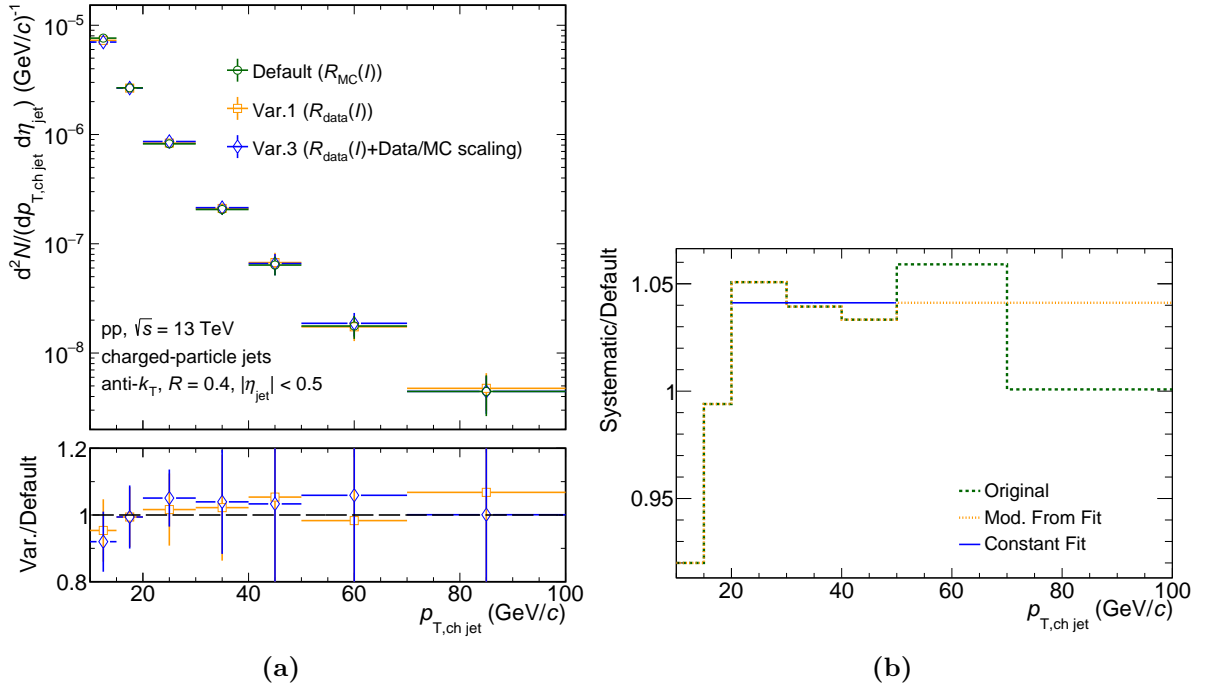

Figure 5.27. – **a)** Comparison of the performance-corrected b-jet spectra for the default template-fitting procedure, Var.1 and Var.3. **b)** Ratio of the b-jet spectrum that is obtained from Var.3 over the default spectrum. The fit for the extrapolation of the systematic uncertainty to large $p_{\text{T,ch jet}}$ is also displayed.

the description of the total and p_T-dependent abundance of the V^0 particles K_s^0, Λ^0 or $\overline{\Lambda}^0$ in comparison to data from ALICE and CMS [20, 148]. Jets containing these particles depict a considerable background source for this analysis: tracks that originate from V^0 decays in general have a large Sd_{xy} as they exhibit long lifetimes – about $c\tau_{K_s^0} = 2.88\,\mathrm{cm}$ for K_s^0 mesons and about $c\tau_\Lambda = 7.89\,\mathrm{cm}$ for Λ baryons [160]. This is why, as a systematic variation, a separate *V^0-jet template* is obtained via a dedicated V^0-jet tagger from the real data. Thereby, a jet is called a V^0 jet if the constituent with the largest Sd_{xy} is a *V^0-daughter track*, meaning a particle that originates from a V^0 decay. The modified performance estimation is done by template fits with four independent fit templates for uds, c, b and V^0 jets and the tagged b-jet spectrum will be corrected with the resulting performance observables.

Obtaining the V^0-jet template

As a first step, tracks in the data are tagged as V^0 candidates via the cut selection provided in Tab. 5.3, which is based on the corresponding selection in previous analyses [18, 149]. For detailed information on the cut selection, it is referred to these sources. To decrease the bias on the $\ln(\mathrm{JP})$ distributions that would arise from a cut on the transverse impact-parameter d_0, the respective cut for the V^0 daughter tracks is adjusted to $d_0 > 0.05\,\mathrm{cm}$. It can not be opened further due to a fixed selection criterion in the ALICE pre-selection of V^0 candidates.

If the N1-type track of a jet object is identified as a V^0-daughter track, the respective jet is called a V^0 jet. The raw V^0-tagged distributions are stored in dependence of $\ln(\mathrm{JP})$ and $p_{\mathrm{T,ch\ jet}}$ for further analysis. To avoid double counting, the fit templates for c and light-flavour jets are obtained by rejecting V^0-jet candidates, while beauty jets that contain a V^0-decay track are classified as b jets.

Similar as for the default template-fitting procedure, data toy studies lead to the final result for the b-jet fraction and its uncertainty in the b-tagged and the b-untagged sample. However, in addition to the data, also 300 toy samples are generated for the V^0-jet template. Thus, every data-toy distribution is fit by the sum of the original b, c and light-flavour templates and a V^0-jet toy distribution. In this way, the significant statistical uncertainties of the V^0-jet templates are captured by the statistical uncertainties of the results for the b-jet fractions. The central values for the b-jet fraction of one specific fit range as well as the corresponding uncertainties are calculated as averages from the individual fit results.

For each of the V^0-jet toys, every bin in $p_{\mathrm{T,ch\ jet}}$ and $\ln(\mathrm{JP})$ of the raw V^0-tagged jets is varied according to a Poisson distribution with the mean set to the original bin content. Every toy is then corrected for the efficiency and purity of the V^0-jet selection as a function of $p_{\mathrm{T,ch\ jet}}$ and $\ln(\mathrm{JP})$ based on the LHC18f5 MC production. In this way, the input from PYTHIA 8 simulations is necessary for the design of the V^0-jet tagger, however, the information is reduced to the description of the V^0 decay kinematics in dependence of $p_{\mathrm{T,ch\ jet}}$ and $\ln(\mathrm{JP})$. The latter is assumed to be captured by PYTHIA 8 simulations with sufficient precision.

The purity

$$p_{V^0} = \frac{N_{V^0}^{\mathrm{True\ Tag}}}{N_{V^0}^{\mathrm{Tag}}} \tag{5.6.6}$$

is defined as the number of true V^0-jet tags $N_{V^0}^{\mathrm{True\ Tag}}$ divided by the number of tagged V^0 jets $N_{V^0}^{\mathrm{Tag}}$ and the efficiency ϵ_{V^0} is given by

$$\epsilon_{V^0} = \frac{N_{V^0}^{\mathrm{Tag}}}{N_{V^0}} \tag{5.6.7}$$

with the number of true V^0 jets N_{V^0} in the LHC18f5 MC production. The performance correction is done separately for the b-tagged and the b-untagged data samples as well as for K_s^0 and Λ particles and it is applied on a per-jet basis. This means that each jet that has been selected as a V^0-jet candidate is corrected by a factor p_{V^0}/ϵ_{V^0} depending on its $p_{\mathrm{T,ch\ jet}}$ and $\ln(\mathrm{JP})$. Due

Table 5.3. – Cuts for the selection of V^0 daughter tracks as inspired from [18, 149]. Cuts that have been chosen differently than the cited analysis are marked with *.

V^0 candidates	
Online or On-The-Fly	offline
d_0 (cm)*	> 0.05
DCA Pos to Neg (σ)	1
$\cos(\theta_{PA})$ for K_s^0	> 0.97
$\cos(\theta_{PA})$ for Λ	> 0.995
y	< 0.5
V^0 decay radius (cm)	> 0.5
proper lifetime K_s^0 (cm)	$7.5\,c\tau_{K_s^0}$
proper lifetime Λ (cm)	$4\,c\tau_\Lambda$
Competing V^0 rejection K_s^0 (GeV/c)	< 0.005
Competing V^0 rejection Λ (GeV/c)	< 0.010
Armenteros-Podolanski cut*	$p_{T,Arm} > 0.2\alpha_{Arm}$
V^0 daughter tracks	
dE/dx (σ)	3
TPC refit required	true
Kink candidate	false
number of TPC clusters	> 0
number of TPC crossed rows	> 70
TPC crossed rows / TPC clusters	> 0.8
$|\eta|$	< 0.8

to the $d_0 > 0.05$ cut for V^0-decay particles, the Sd_{xy} distributions for N1-type tracks suffer from threshold effects around $Sd_{xy} \approx 0$. This is why, a cut of $Sd_{xy} > 3$ is applied for $N_{V^0}^{\text{Tag}}$ and $N_{V^0}^{\text{True Tag}}$ in Eqn. 5.6.6 and Eqn. 5.6.7. The latter is corrected for by the efficiency correction in Eqn. 5.6.7 as N_{V^0} is defined for $Sd_{xy} > 0$.

The resulting two-dimensional efficiencies and purities for the V^0-jet tagger are displayed in Fig. 5.28. Because of the low statistics at large ln(JP), the bin contents are averaged over several $p_{T,\text{ch jet}}$ bins in this regime. In the case of jets containing a K_s^0-decay particle, the bin contents for $20 < p_{T,\text{ch jet}} < 50\,\text{GeV}/c$ and $50 < p_{T,\text{ch jet}} < 100\,\text{GeV}/c$ are averaged for $-\ln(\text{JP}) > 22$. In the case of jets containing a Λ particle, the bin contents for $5 < p_{T,\text{ch jet}} < 20\,\text{GeV}/c$ and $20 < p_{T,\text{ch jet}} < 100\,\text{GeV}/c$ are averaged for $-\ln(\text{JP}) > 19$. Jets with $-\ln(\text{JP}) > 22$ are rejected if they contain Λ particles as otherwise, the performance correction is suffering from outliers due to the low statistics.

To test the reliability of the performance correction described above, the LHC18f5 MC production is split into two equally-sized samples of which one is used for the generation of the purities and efficiencies whereas the other one is utilised to test the MC closure. In this sense, the performance-corrected V^0-jet template that is obtained for the testing sample is compared to the respective MC truth distribution. The results of this MC closure test for $10 < p_{T,\text{ch jet}} < 15\,\text{GeV}/c$ are provided exemplary in Fig. 5.29 for the b-tagged and the b-untagged test sample. Apart from statistical fluctuations, the respective ratios are compatible with one.

The final V^0-jet ln(JP) templates obtained from the real data are compared to the ln(JP) distributions from the LHC18f5 MC production in Fig. 5.30 for b-tagged and b-untagged jets with $10 < p_{T,\text{ch jet}} < 15\,\text{GeV}/c$. Although no quantitative statements can be made due to the considerable statistical uncertainties, the MC simulations appear to be more peaked and shifted towards large Sd_{xy} values.
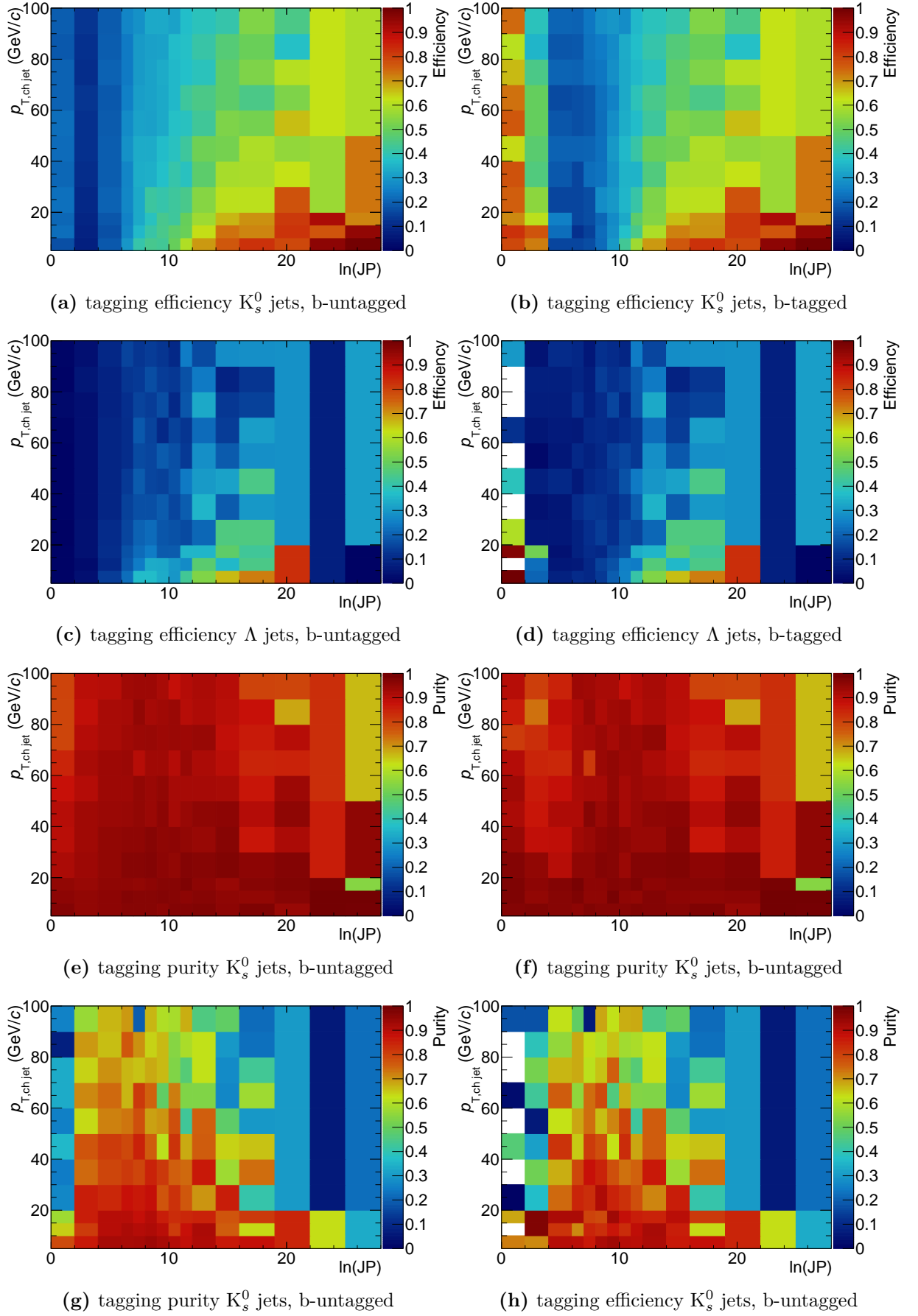
(a) tagging efficiency K_s^0 jets, b-untagged

(b) tagging efficiency K_s^0 jets, b-tagged

(c) tagging efficiency Λ jets, b-untagged

(d) tagging efficiency Λ jets, b-tagged

(e) tagging purity K_s^0 jets, b-untagged

(f) tagging purity K_s^0 jets, b-untagged

(g) tagging purity K_s^0 jets, b-untagged

(h) tagging efficiency K_s^0 jets, b-untagged

Figure 5.28. – The two-dimensional efficiencies and purities for the selection of jets that contain K_s^0- and Λ-decay particles. The tagging is determined separately for b-untagged (left column) and b-tagged (right column) jets.
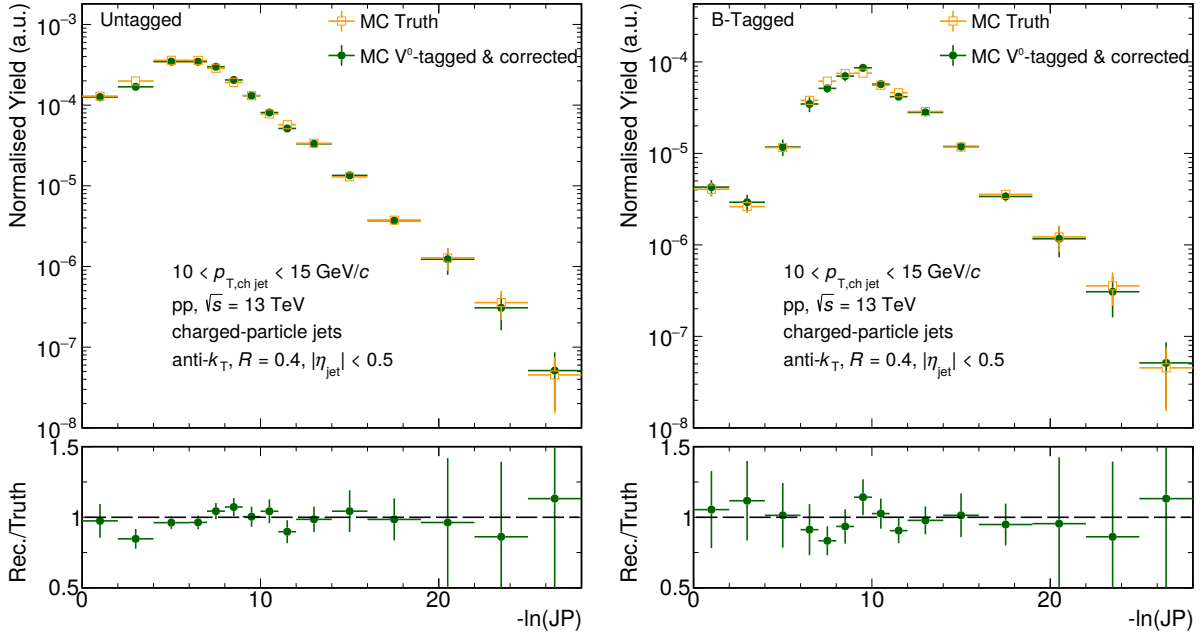
Figure 5.29. – MC Closure test: The ln(JP) distributions for V^0 jets from MC truth information compared to the ones that have been obtained using the V^0-jet tagger for $10 < p_{\text{T,ch jet}} < 15\,\text{GeV}/c$. The distributions are provided for b-untagged (left) and b-tagged (right) jets. The bottom panel shows the ratio of the spectrum obtained by the V^0-jet tagger ("Rec.") over the one from MC truth information ("Truth").
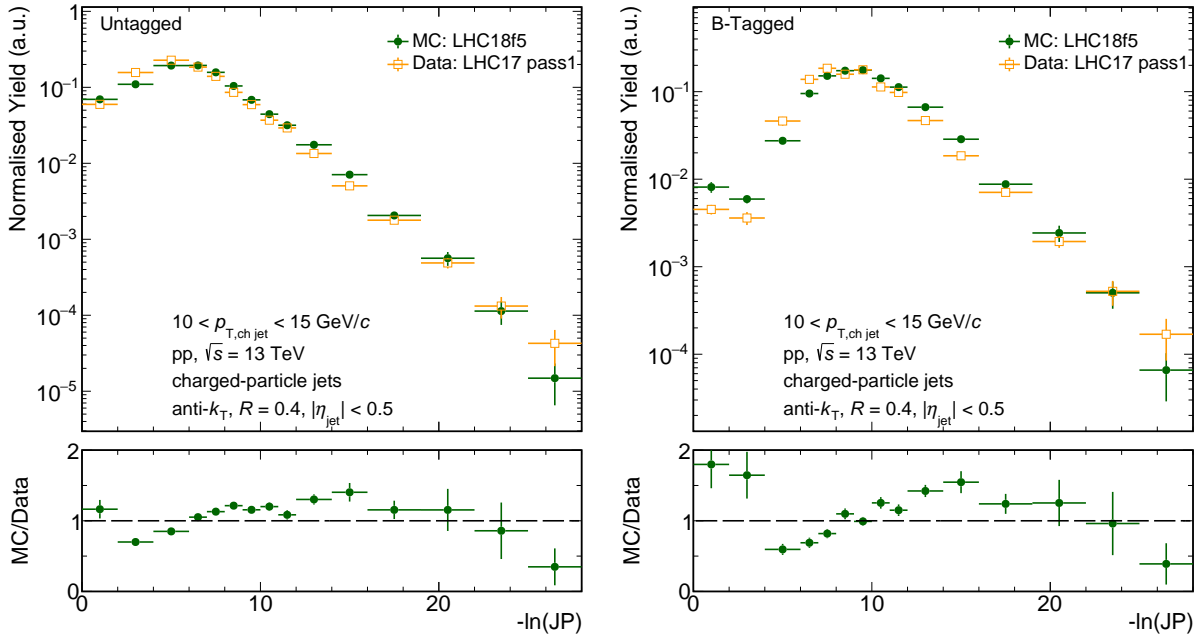

Figure 5.30. – Comparison of the V^0-tagged and performance-corrected ln(JP) distributions from the real data and the corresponding distributions from MC truth detector-level information for the b-untagged (left) and the b-tagged (right) data sample.
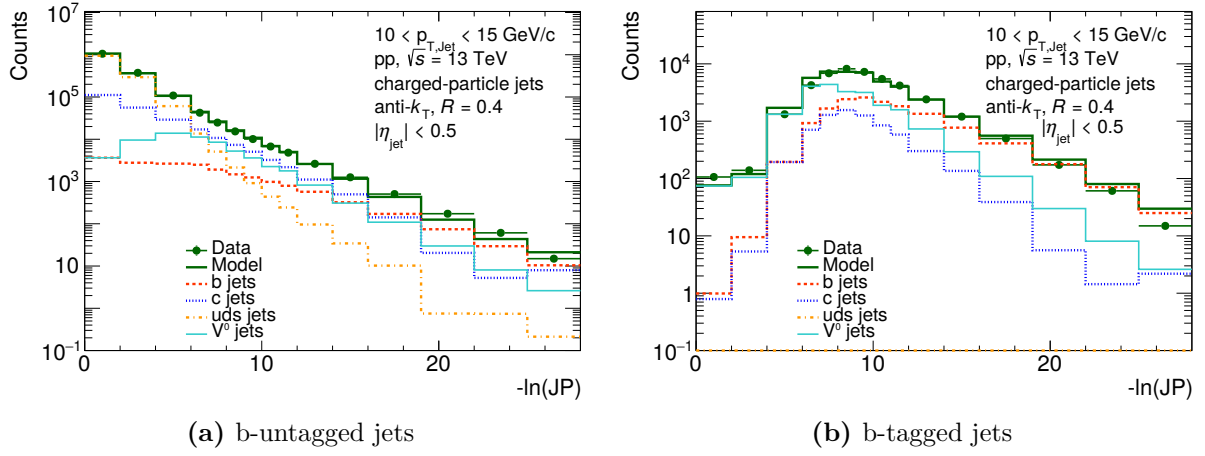
(a) b-untagged jets   **(b)** b-tagged jets

Figure 5.31. – Template fits to b-untagged (left) and b-tagged jets (right) with $10 < p_{\text{T,ch jet}} < 15\,\text{GeV}/c$ for the performance estimation of the b-jet tagger using an additional fourth template for V^0-jets. The contributions of the individual jet flavours to the fit model are scaled with respect to their relative contribution to the fit model.

Template-fit results with additional V^0-jet template

The V^0-jet ln(JP) distributions which have been obtained by the dedicated tagger are fed to the procedure for the performance estimation of the b-jet tagger. Thereby, the contribution of the V^0-jet templates is fixed to the abundance obtained from the data and only the contributions from b, c and light-flavour jets are free parameters of the template fits. In comparison to the default procedure, the binning in ln(JP) is chosen to be wider at low ln(JP) due to the low statistics in this regime. Still, it is not feasible with the current statistics to perform the fitting for $p_{\text{T,ch jet}} > 40\,\text{GeV}/c$. This is why the systematic uncertainties are extrapolated from low to large $p_{\text{T,ch jet}}$, as detailed later on.

The results for the final b-jet fraction within one specific $p_{\text{T,ch jet}}$ bin are obtained from averaging the results for different fit ranges, as it has been done for the default fitting procedure. Due to the modified binning at low ln(JP), the plateau region is defined differently than for the default procedure:

- b-tagged: $0 < -\ln(\text{JP}) < 5$,

- b-untagged ($p_{\text{T,ch jet}} \leq 20\,\text{GeV}/c$): $4 < -\ln(\text{JP}) < 6$,

- b-untagged ($p_{\text{T,ch jet}} > 20\,\text{GeV}/c$): $2 < -\ln(\text{JP}) < 5$.

In Fig. 5.31, the template fits for the default fit range and $10 < p_{\text{T,ch jet}} < 15\,\text{GeV}/c$ are displayed exemplarily. It can be observed, that V^0 jets contribute significantly at low $p_{\text{T,ch jet}}$, where they dominate the region of $-\ln(\text{JP}) \leq 6$ in the b-tagged data. In comparison to the fit template for light flavours, the templates for V^0 jets reach towards larger $-\ln(\text{JP})$ and, in case of the b-untagged distribution, exhibit a dip at low $-\ln(\text{JP})$. The results from template fits for the remaining $p_{\text{T,ch jet}}$ intervals show that the V^0-jet contribution is decreasing with increasing $p_{\text{T,ch jet}}$ (see Fig. C.4).

In Fig. 5.32 and Fig. 5.33a, the results for the b-jet fractions in the b-tagged and the b-untagged data sample, the efficiency and the resulting performance-corrected b-jet spectrum are compared to the results for the default performance correction. Clearly, the statistical uncertainties are larger for the observables obtained from the systematic variation than those of the default correction procedure. Considering this, the results for the b-jet spectrum from the default and the modified procedure are compatible with each other. Having a closer look, the fraction for b-tagged jets with $10 < p_{\text{T,ch jet}} < 15\,\text{GeV}/c$ is significantly larger by about $20\,\%$ for the modified
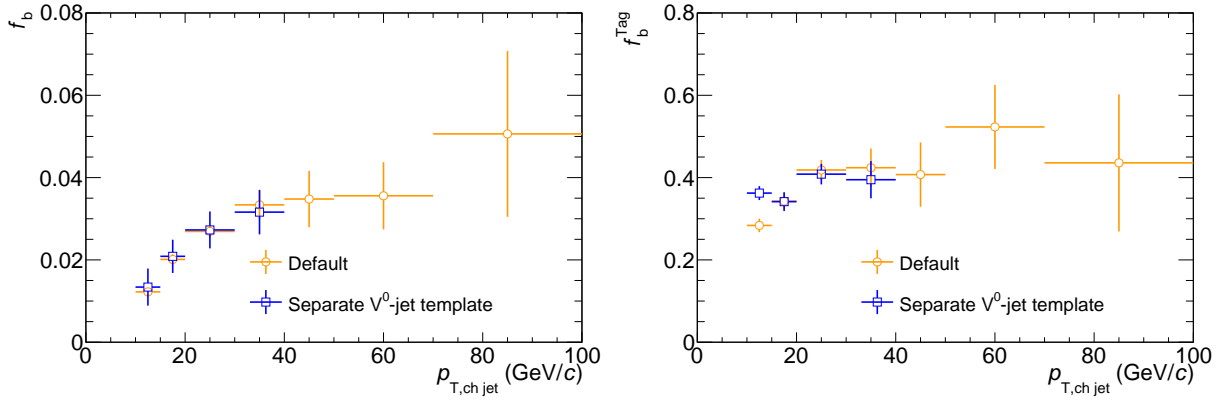
Figure 5.32. – Comparison of the b-jet fractions in the b-untagged (left) and b-tagged (right) data sample for the default method of performance correction as well as for using a separate dedicated V^0-jet fit template.

performance correction. All other fractions are, however, in good agreement for the default and the modified procedure.

As it was said before, the template fits have only been performed for $p_{\text{T,ch jet}} < 40\,\text{GeV}/c$ due to the limited statistics of the V^0-jet templates for large $p_{\text{T,ch jet}}$. Therefore, the relative systematic uncertainty for $30 < p_{\text{T,ch jet}} < 40\,\text{GeV}/c$ is adopted also for $p_{\text{T,ch jet}} > 40\,\text{GeV}/c$ as shown in Fig. 5.33b. Since the V^0-jet contribution is expected to decrease with increasing $p_{\text{T,ch jet}}$, this assumption most probably leads to an overestimation of the corresponding uncertainties.

A note on prospects for further analysis   The results of the method for the b-jet tagging performance estimation discussed in this section will gain in precision with increasing statistics. They will thus profit from including the full available ALICE data for 13 TeV pp collisions. The selection criteria that require the V^0-daughter tracks to have $d_0 > 0.05$ as well as $|\eta| < 0.8$, have originally been optimised for analyses that demand for high purities of the V^0-tagged samples. Since b-jet analyses solely aim at rejecting jets containing V^0-daughter tracks, thus gaining advantage from a higher efficiency not purity, the above-mentioned selections could be investigated in more detail in subsequent analyses.

To conclude, since V^0 jets appear to contribute significantly to the region of large ln(JP) for low jet momenta, the close connection of b-jet and V^0 tagging will be beneficial for future measurements that follow a similar analysis concept as it has been applied for this thesis and aim for sensible precision at low $p_{\text{T,ch jet}}$.

5.6.3. Further uncertainties on the performance-corrected spectrum

Variation of the plateau region for the template fits   As described in Sect. 5.4.2, the results for the b-jet fractions of the template fitting procedure are averaged over the results that are obtained for different fit ranges. The corresponding interval of the lower fit boundary for which the results have been averaged has been called the plateau region. For the estimation of the systematic uncertainties with respect to the choice of the plateau region, the upper limit and the lower limit of the plateau region are varied independently for the b-tagged data as well as for the untagged data for $p_{\text{T,ch jet}} \leq 20\,\text{GeV}/c$ and for the untagged data for $p_{\text{T,ch jet}} > 20\,\text{GeV}/c$, as listed in Tab. 5.4. The following combinations of variations have been performed:

- Tagged Low: the plateau region for the b-tagged data sample has been varied according to variation "Low".

- Tagged Low, Untag Low: the plateau region for the b-tagged and the b-untagged data sample has been varied according to variation "Low".

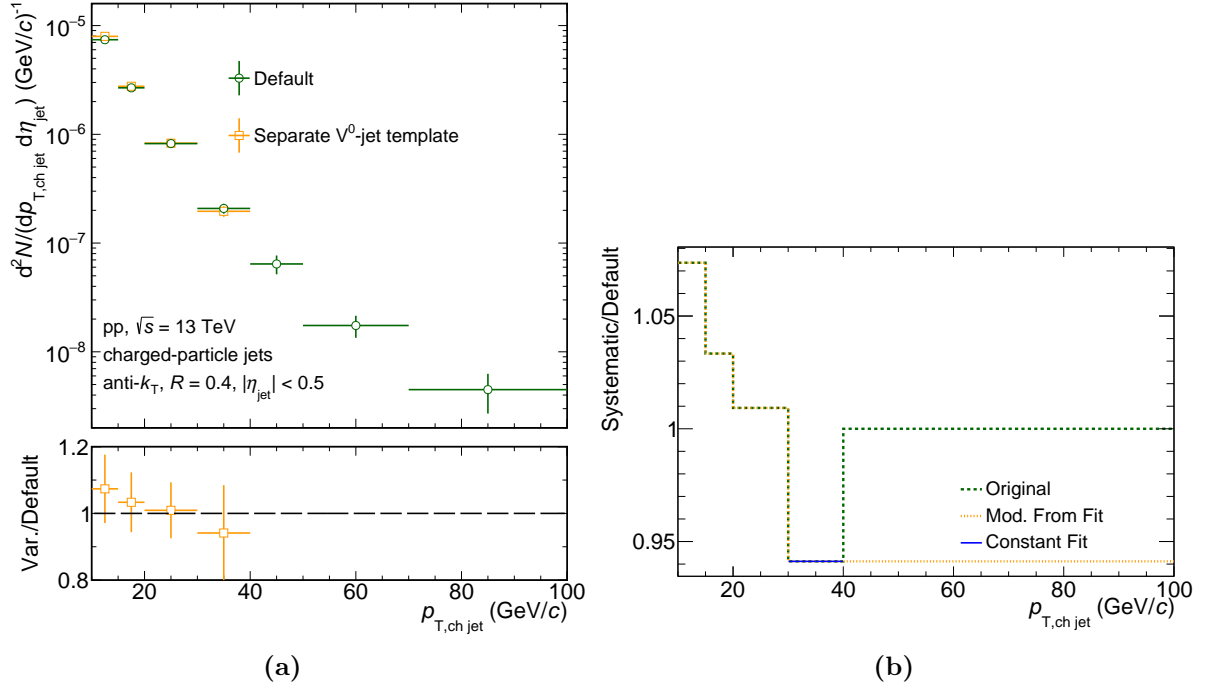**(a)**   **(b)**

Figure 5.33. – **a)** Comparison of the performance-corrected b-jet spectra for the default template-fitting procedure and for the systematic variation with respect to an additional V^0-jet template. **b)** Ratio of the b-jet spectrum that is obtained via the consideration of a dedicated V^0-jet template over the default spectrum. The fit for the extrapolation of the systematic uncertainty to large $p_{\mathrm{T,ch\ jet}}$ is also displayed.

Table 5.4. – The variations with respect to the interval of the lower boundary $\ln(\mathrm{JP})_{\mathrm{min}}$ of the fit range over which the fit results for the b-jet fraction in the b-tagged and the untagged data sample have been averaged.

	Default	High	Low
Tagged	$0<\ln(\mathrm{JP})<6$		$0<\ln(\mathrm{JP})<4$
Untagged	$2<\ln(\mathrm{JP})<7$	$4<\ln(\mathrm{JP})<7$	$2<\ln(\mathrm{JP})<5$
($p_{\mathrm{T,ch\ jet}}<20\,\mathrm{GeV/c}$)			
Untagged	$1<\ln(\mathrm{JP})<6$	$2<\ln(\mathrm{JP})<6$	$1<\ln(\mathrm{JP})<4$
($p_{\mathrm{T,ch\ jet}}>20\,\mathrm{GeV/c}$)			

- Tagged Low, Untagged High: the plateau region for the b-tagged data sample has been varied according to variation "Low" and the one for the b-untagged data sample according to variation "High".

- Untagged Low: the plateau region for the b-untagged data sample has been varied according to variation "Low".

- Untagged High: the plateau region for the b-untagged data sample has been varied according to variation "High".

For the b-tagged data, no shift of the plateau region to larger values of the lower fit boundary has been investigated as for this sample, the fit results are much more stable for different fit ranges and the systematic effect is expected to be negligible compared to the listed variations.

Variation of the tagger working point   For the default settings of the b-jet tagger, $p_{\mathrm{T,ch\ jet}}$-dependent threshold cuts are applied simultaneously to N1- and N2-type tracks as described
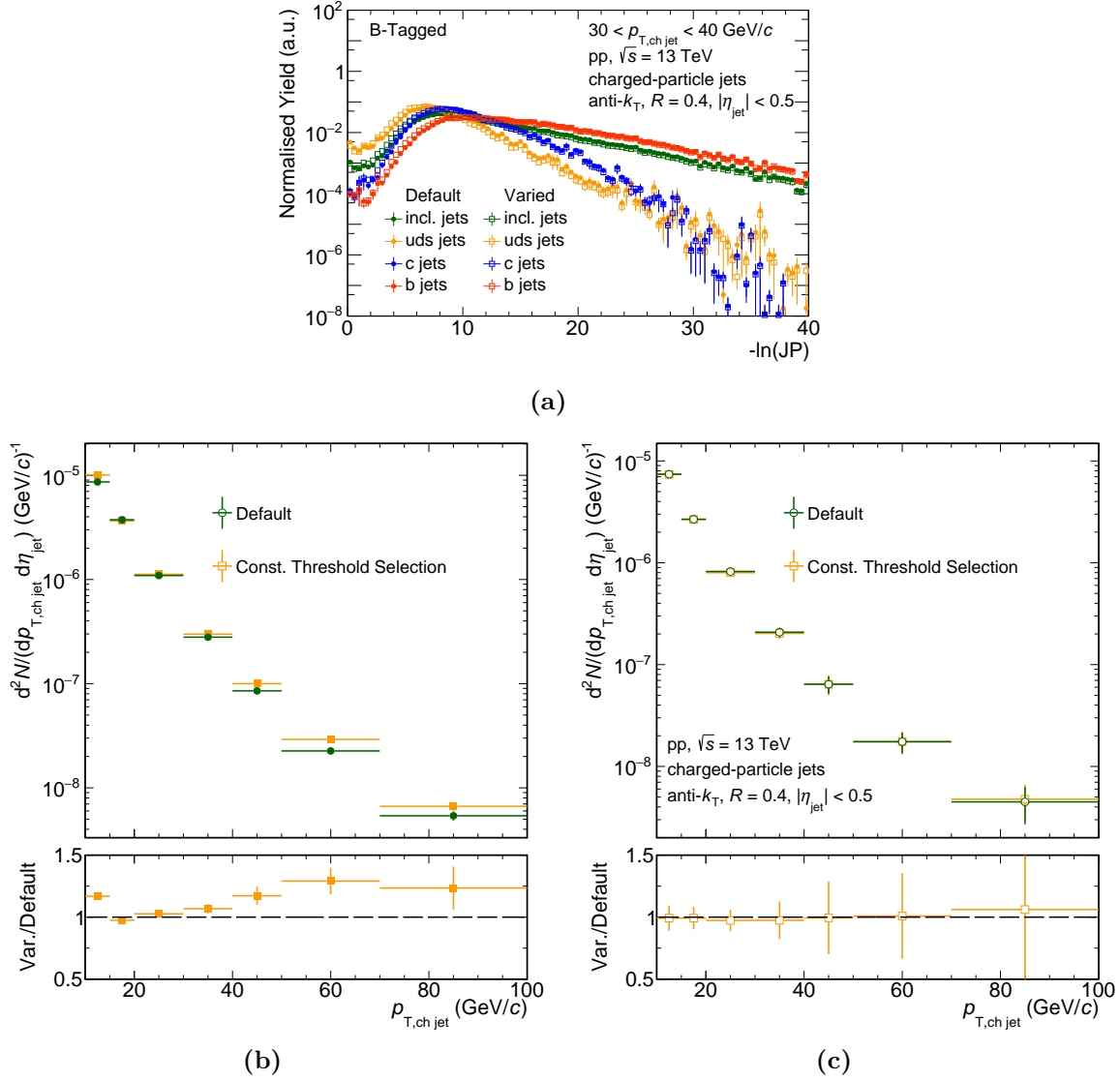
(a)

(b)

(c)

Figure 5.34. – Comparison of different distributions for the default b-jet tagger setting and for threshold selections with cuts that are constant over the full $p_{\mathrm{T,ch\,jet}}$ range. **a)** the ln(JP) templates for the performance estimation for b-tagged jets with $30 < p_{\mathrm{T,ch\,jet}} < 40\,\mathrm{GeV}/c$. **b)** the raw tagged b-jet spectra before the performance correction. **c)** the performance-corrected b-jet spectra.

in Sect. 5.3. As a systematic variation, a threshold selection with cuts that are constant with $p_{\mathrm{T,ch\,jet}}$ is applied, whereby the selection criteria are set to $Sd_{xy}^{\mathrm{Thresh}} = 4$ for N1- and $Sd_{xy}^{\mathrm{Thresh}} = 2$ for N2-type tracks.

To estimate the performance of this modified tagger, the ln(JP) templates for b-tagged jets need to be adjusted accordingly. As it can be seen in Fig. 5.34a, the corresponding distributions exhibit a slight shift towards larger $-\ln(\mathrm{JP})$ values with respect to the default distributions. The raw tagged b-jet spectra before the performance correction exhibit significant deviations of up to about 25 % for the default and modified selection criteria (see Fig. 5.34b). This difference is eliminated by the performance correction as it can be seen in Fig. 5.34c, where the performance-corrected b-jet spectra are compared for the default and the modified tagger setting. These distributions are compatible within the statistical uncertainties, which confirms the correct functioning of the procedure that is utilised for the performance correction.

Variation of the track selection In Sect. 5.2.2, it has been described that the b-jet selection is performed on the basis of tracks which have been reconstructed with high quality. In particular, only tracks with a number of TPC clusters $n_{\mathrm{Cl}}^{\mathrm{TPC}} > 100$ are used for the jet finding (see Tab.5.2). The choice of appropriate selection criteria is not explicitly defined, which is why two systematic variations are considered with respect to this cut:

1) Modified $n_{\mathrm{Cl}}^{\mathrm{TPC}}$: tracks with $n_{\mathrm{Cl}}^{\mathrm{TPC}} > 80$ are accepted,

2) Crossed rows: tracks with $n_{\mathrm{Cl}}^{\mathrm{TPC}} > 0$, a number of TPC crossed rows over findable clusters greater than 0.8 and a number of TPC crossed rows greater than 70 are accepted.

In principle, the following steps need to be performed to obtain the b-jet spectrum for the modified track-selection criteria:

a) the resolution function $R_{\mathrm{MC}}(I)$ which is necessary for the ln(JP) calculation needs to be re-determined for all track and jet categories listed in Sect. 5.4.1,

b) the ln(JP) distributions for the template-fitting procedure need to be re-determined,

c) the template fits need to be performed using the modified ln(JP) distributions for the templates and the data,

d) the performance-corrected b-jet spectra need to be unfolded using a dedicated response matrix which has been obtained using the modified track selection criteria.

The need for step d) in the upper listing identifies the variation with respect to the track selection as a category 2 uncertainty, as defined at the beginning of this section.
It has been found that differences of the resolution functions for the default and the modified track selections are negligible. This is why, step a) is only performed exemplarily for the variations "Modified $n_{\mathrm{Cl}}^{\mathrm{TPC}}$", while for the variations "Crossed rows", the resolution functions of the default analysis have been utilised. The modified track selection criteria in the latter case still affect the distribution of the intrinsic impact parameter I_0 that is fed to Eqn. 5.4.1 as well as the Sd_{xy} distributions of N1- and N2-type tracks based on which the b-jet selection criteria are formulated. Steps b) to d) are implemented for both variations listed above.

The modified ln(JP) templates for b-tagged and untagged jets with $30 < p_{\mathrm{T,ch\ jet}} < 40\,\mathrm{GeV}/c$ for the variation "Modified $n_{\mathrm{Cl}}^{\mathrm{TPC}}$" are provided exemplarily in Fig. 5.35. There are no significant differences of the default and the modified distributions. Fig. 5.36 compares the b-jet spectra before and after the performance correction for the default and the modified track selection. The differences with respect to the performance-corrected spectrum are also negligible within the statistical uncertainties.

MC-template statistics The influence of the statistics of the ln(JP) fit templates that are used for the performance estimation of the b tagger is investigated as a systematic uncertainty. As it has been described in Sect. 5.4.3, the central value for the b-jet fractions in the b-tagged and the untagged data are obtained from data toy studies. Thereby, toy data sets are generated from the original data by varying the bin contents according to a Poisson distribution. To investigate the effect of the statistics of the MC ln(JP) templates, toys are also generated from the fit templates. This is done by varying the bin content of the fit templates according to a Gaussian distribution with the mean set to the original bin content. Thus, every data toy is fit with a sum of toys for the ln(JP) fit templates. The b-jet fractions for one fit range and $p_{\mathrm{T,ch\ jet}}$ bin are obtained by averaging over the results of the individual fits.
The deviations of the performance-corrected b-jet spectra for the default method and for considering the statistics of the MC fit templates are negligible (see Fig. 5.37).
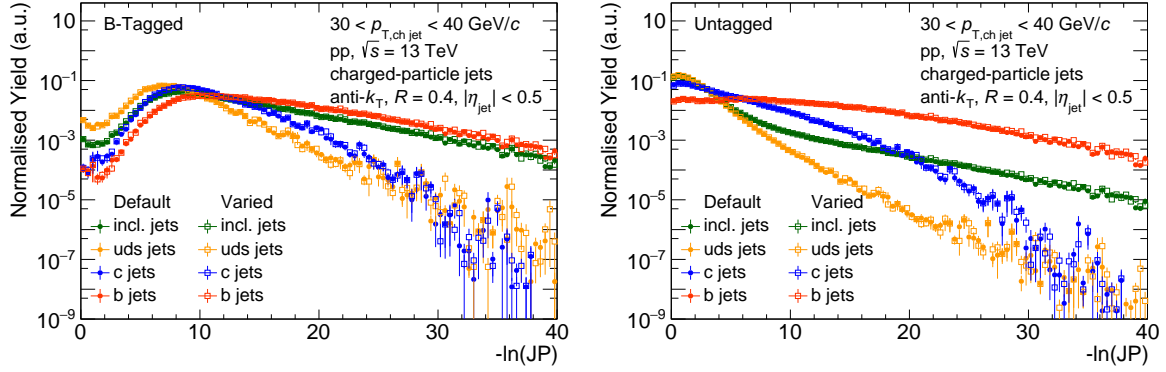
Figure 5.35. – The ln(JP) templates for the default track selection and variation "Modified $n_{\mathrm{Cl}}^{\mathrm{TPC}}$" are compared for different jet flavours. The distributions are shown for $30 < p_{\mathrm{T,ch\ jet}} < 40\,\mathrm{GeV}/c$ for b-tagged (left) jets and untagged jets (right).
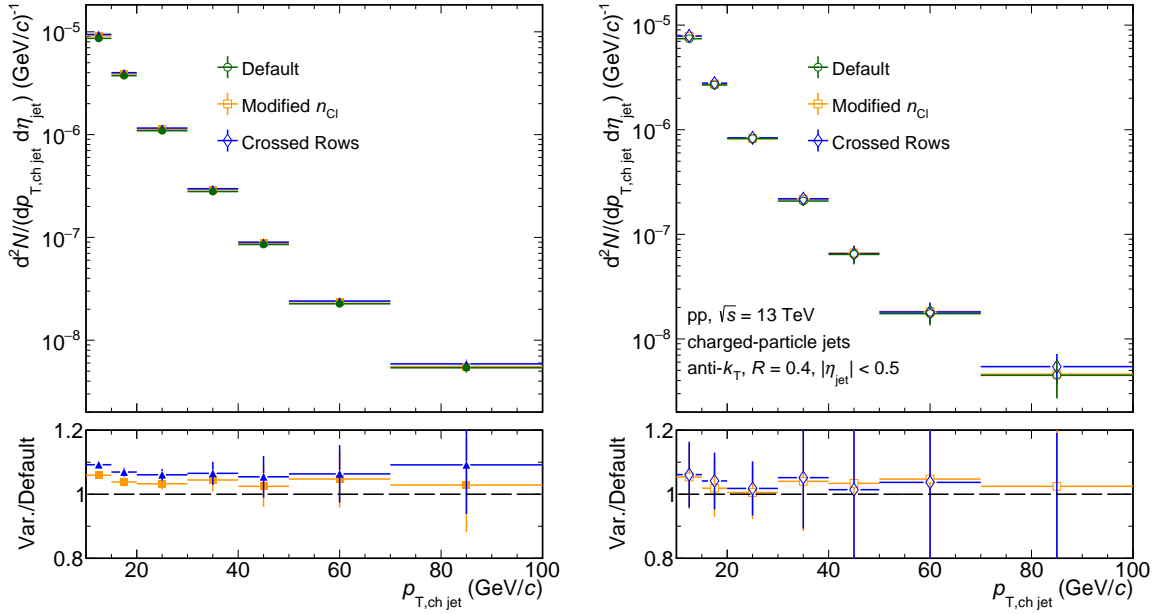

Figure 5.36. – Comparison of the b-jet spectra for the default and the modified track selection. **Left:** the raw tagged b-jet spectra before the performance correction. **Right:** the performance-corrected b-jet spectra.
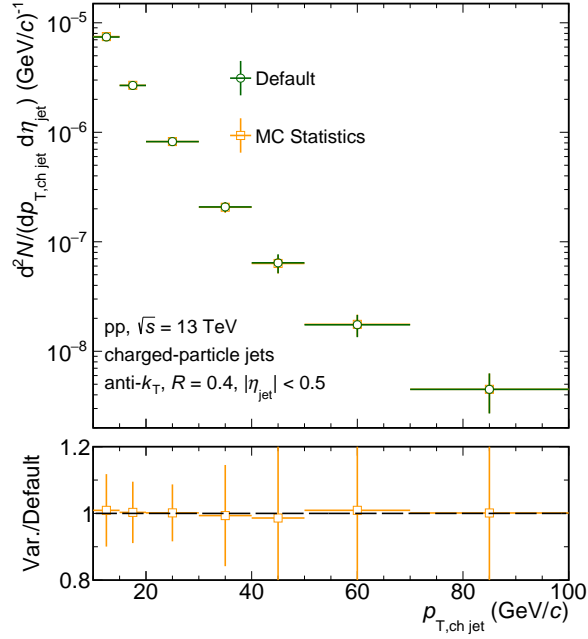
Figure 5.37. – Comparison of performance-corrected b-jet spectra for the default performance correction and for considering the statistics of the the ln(JP) fit templates from MC simulations.

5.6.4. Unfolding

Unfolding procedures can show sensitivity with regards to several aspects like the steepness or statistical fluctuations of the smeared data as well as the choice of the prior distribution. For this analysis, the following variations of the default unfolding method have been investigated to identify potential dependencies and assign corresponding systematic uncertainties:

- **truncation uncertainty:** the lower boundary of the first bin of the measured data is changed from $10\,\mathrm{GeV}/c$ to $5\,\mathrm{GeV}/c$. The raw number of tagged b jets in this first bin is corrected by the efficiency and purity for the wider bin using the default method for the performance correction.

- **regularisation uncertainty**: the number of iterations is changed from the default value $\beta = 4$ to $\beta = 3$ and 5.

- **prior uncertainty:** the default PYTHIA 8 prior is replaced by the POWHEG+PYTHIA prior which has already been used for the MC stability tests of the default method in Sect. 5.5.3.

- **inner-binning uncertainty:** the smeared data is rebinned such that all intermediate bin boundaries are shifted by $\pm 2\,\mathrm{GeV}/c$ and the lower and upper boundaries of the spectrum are untouched. For more details see below.

- **uncertainty wrt. choice of algorithm:** instead of Iterative Bayesian unfolding, the method of SVD unfolding [106, 111] is applied. For more details see below.

- **tracking efficiency uncertainty:** a modified response matrix is obtained via randomly rejecting $4\,\%$ of the detector-level tracks. The difference of the unfolded spectrum as it is obtained using the default response matrix and the modified response matrix is taken as systematic uncertainty. Strictly speaking, this uncertainty does not arise from the unfolding procedure but it can easily be implemented by using the latter. This is why, this uncertainty is also addressed in this section.
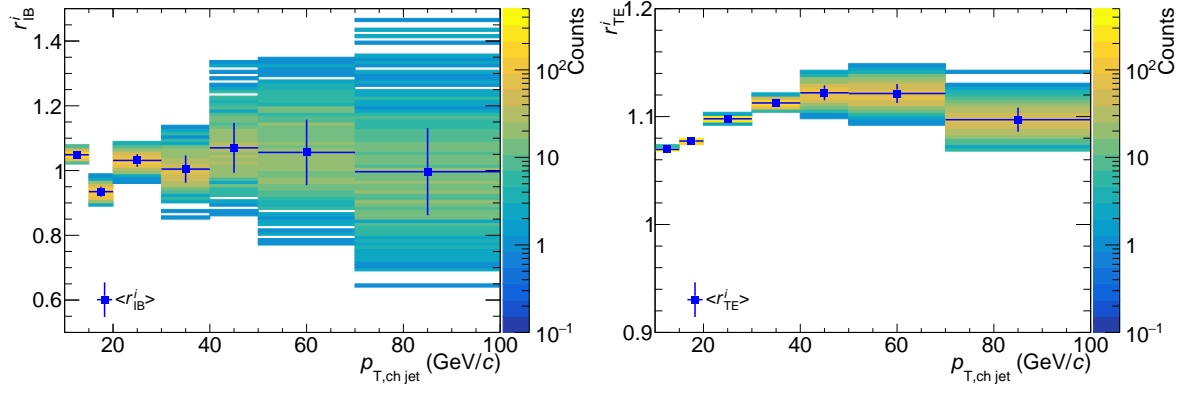
Figure 5.38. – The spread of the ratios r_v^i for toys $i \in [1, \ldots, 500]$ for the procedures to estimate the inner-binning uncertainty (left, $v = \text{IB}$) and the tracking efficiency uncertainty (right, $v = \text{TE}$).

For each of these categories of unfolding uncertainties, the maximum deviation will be taken as resulting uncertainty. The respective resulting uncertainties are added in quadrature to obtain the final unfolding systematic uncertainty Δ_{unfold}.

For the default unfolding procedure, toys of the smeared data are obtained, which are then varied within their statistical uncertainties. Every toy is unfolded and the mean of the individual unfolded toys is adopted as the final unfolded b-jet spectrum. To reduce the effect of the statistical uncertainties on the systematics uncertainties, this procedure is modified for the estimation of the latter. The bin contents of the raw b-jet spectrum before the performance correction are varied according to a Poisson distribution and 500 toy distributions are generated. Thereby, the mean of the Poisson distribution is set to the respective original bin content. After this, the toys are corrected for the performance and every toy is unfolded with the default settings *and* the settings that are investigated as systematic variation. Thus, for every toy $i \in [1, \ldots, 500]$, an unfolded distribution $t_{\text{def}}^i(p_{\text{T,ch jet}})$ is obtained from the default unfolding procedure and an unfolded distribution $t_v^i(p_{\text{T,ch jet}})$ is obtained from the varied procedure v. The mean of the ratios of the latter distributions

$$\Delta_{\text{unfold}}^v(p_{\text{T,ch jet}}) = \left\langle \underbrace{\frac{t_v^i(p_{\text{T,ch jet}})}{t_{\text{def}}^i(p_{\text{T,ch jet}})}}_{r_v^i} \right\rangle \tag{5.6.8}$$

are adopted as systematic uncertainty for variation v. Exemplarily, the spread and the mean of the ratios r_v^i for the estimation of the inner-binning uncertainty and the tracking efficiency uncertainty are displayed in Fig. 5.38.

The determination of the inner-binning uncertainty and of the uncertaintiy with respect to the choice of the unfolding algorithm are more complex, which is why further details will be given in this respect in the next paragraphs.

Inner-Binning Uncertainty

For the estimation of the inner-binning uncertainty, a performance-corrected b-jet spectrum with modified binning is unfolded with the default unfolding procedure. This spectrum could be obtained by rebinning the raw b-tagged spectrum and correcting it with the efficiency and purity that are estimated for the modified binning. However, the efficiency and the purity that have been determined by the template fitting procedure exhibit significant statistical fluctuations.
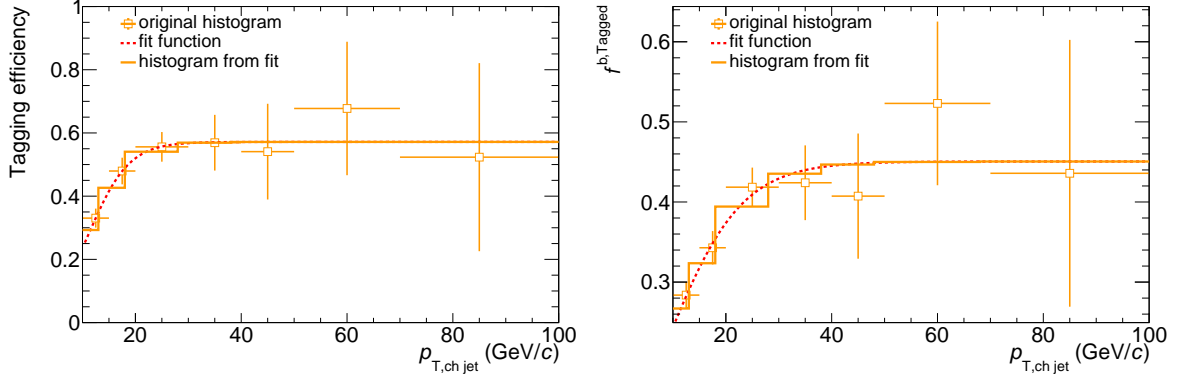
Figure 5.39. – The efficiency (left) and the purity (right) which have been obtained by the default analysis are compared to the fit functions that are used for the estimation of the inner-binning uncertainty. Also shown are the histograms that have been obtained from the fit functions for a shift of the inner bin boundaries by -2 GeV/c.

To disentangle the systematic inner-binning uncertainty from the statistical uncertainty, the $p_{\mathrm{T,ch\,jet}}$-dependent efficiency and the purity are fit with a Sigmoid function

$$s(p_{\mathrm{T,ch\,jet}}) = \frac{a}{(1 + \exp(-b\,(p_{\mathrm{T,ch\,jet}} - c)))} \tag{5.6.9}$$

and the performance observables with modified binning are calculated from the fit functions. The efficiency and purity together with the respective fit functions are shown exemplarily for the variation with -2 GeV/c in Fig. 5.39.

The raw b-tagged spectra with the varied and with the default binning are performance-corrected with the respective efficiencies and purities that are obtained from the fit functions. The inner-binning uncertainty is then calculated based on the ratios of the respective unfolded spectra according to Eqn. 5.6.9.

SVD Unfolding

Unfolding algorithms only approximately recover the true particle-level spectra from smeared detector-level ones. Thereby, each algorithm exhibits particular deficiencies, which is why it is expedient to investigate the characteristics of unfolded spectra that have been obtained with distinct unfolding algorithms. In this analysis, a systematic uncertainty is applied with respect to the difference of the unfolded b-jet spectra for Iterative Bayesian and SVD unfolding. The concept of SVD unfolding is significantly different from Bayesian unfolding and will be introduced shortly in the following.

The problem of deriving the true particle level spectrum x from measurement results b can be formulated as

$$Ax = b \tag{5.6.10}$$

with the response matrix A. The SVD unfolding [106, 111] is based on the method of singular value decomposition (SVD) of the $m \times n$ response matrix according to

$$A = USV^T \tag{5.6.11}$$

with the orthogonal $m \times m$ matrix U, the $n \times n$ matrix V and the diagonal $m \times n$ matrix S. The positive diagonal elements $S_{ii} \equiv s_i$ of the matrix S are called *singular values*. Using Eqn. 5.6.11,

the unfolding problem in Eqn. 5.6.10 can be reformulated like

$$USV^T x = b, \tag{5.6.12}$$

$$\Leftrightarrow Sz = d \quad \text{with vectors} \quad z = V^T x \quad \text{and} \quad d = U^T b. \tag{5.6.13}$$

The equations of this equation system are rearranged such that the s_i are ordered from largest to smallest values.

The equation system in Eqn. 5.6.13 could be solved exactly. However, the corresponding results for z would suffer from strong fluctuations due to measurement uncertainties represented by the uncertainties of d. These will be further amplified by potential small values of s_i. Eqn. 5.6.13 is thus manipulated in several ways to regularise the results. The steps that are taken encompass a rescaling of the response matrix A to suppress the influence of bins with low statistics, the equalisation of the measurement uncertainties and the addition of a regularisation term. The regularised solution to the unfolding problem for bin i of the unfolded spectrum is then given by

$$z_i(\tau) = \frac{d_i}{s_i} \cdot \frac{s_i^2}{s_i^2 + \tau} \tag{5.6.14}$$

with a parameter τ that can be chosen according to the needs of the analysis. The last term in Eqn. 5.6.14 suppresses contributions from all $s_i \lesssim \tau$. This is why, τ should be set to the last element s_k for which the corresponding element d_k is still significant. In the following, the index k will be referred to as the *regularisation parameter.*

In the rest of this section, the application of the SVD algorithm for this analysis is outlined. The following $p_{\text{T,ch jet}}$ binning is utilised for the SVD unfolding:

- smeared data: 10,15,20,30,40,50,70,100 (7 bins)

- particle-level truth/unfolded data: 0,10,15,20,30,50,100,250 (7 bins)

For stable results, the number of bins of the unfolded spectrum should not exceed the number of bins of the detector-level spectrum for the SVD unfolding ($m = n$). Consequently, the unfolded b-jet spectrum is obtained with wider binning in comparison to the results from the Iterative Bayesian unfolding.

To evaluate the stability of the SVD unfolding, the same MC stability tests have been performed as for the Bayesian unfolding (see Sect. 5.5.3). The results of these tests can be found in Appendix C.3. The refolding test, which assesses the mathematical consistency of the unfolding of the real data, is shown in Fig. 5.40a. The test is passed within the statistical uncertainties. As the SVD unfolding does not constitute an iterative approach, a convergence test like it has been done for Iterative Bayesian unfolding is not instructive in this context.

The criteria for the choice of the optimal regularisation parameter k for the SVD unfolding are the following:

1. the element d_k of the d vector introduced in Eqn. 5.6.13 is about one; optimally, k is chosen such that d_k is the first element which is approximately one,

2. minimisation of the total uncertainty which is defined as the quadratic sum of the statistical, the regularisation and the prior uncertainty,

3. no visible bin-to-bin correlations of the unfolded spectrum as illustrated by the Pearson matrices.

The behaviour of the elements d_k of the d vector is displayed in Fig. 5.41b in dependence of k. For $k \geq 3$, the elements d_k are approximately unity.
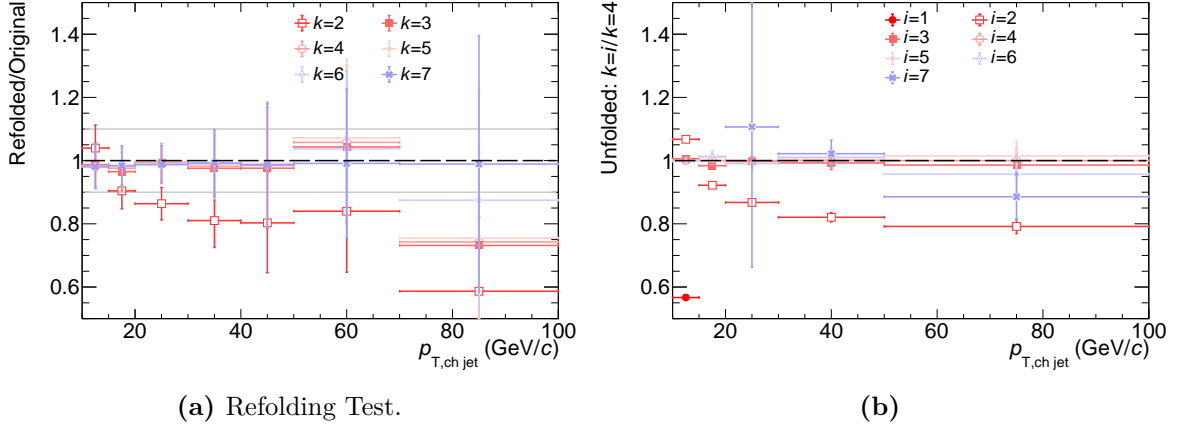
(a) Refolding Test.

(b)

Figure 5.40. – **a)** The ratio of the refolded data over the original performance-corrected data. **b)** The unfolded spectrum for $k = 4$ divided by the unfolded spectra for larger and smaller regularisation parameters.
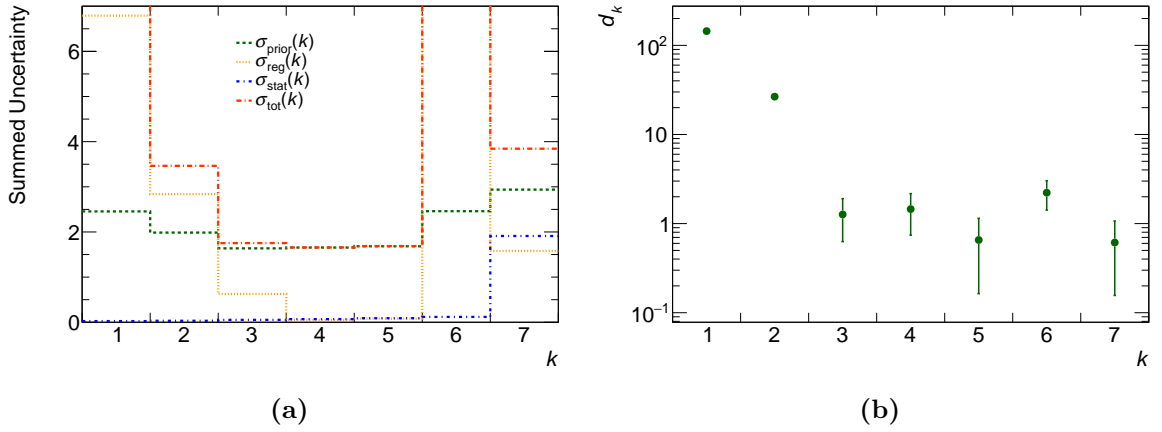

(a)

(b)

Figure 5.41. – **a)** The dependence of the statistical, regularisation and prior uncertainties on the regularisation parameter k. The total unfolding uncertainty is, in this context, given by the quadrature sum of the individual uncertainties. **b)** The elements d_k of the d-vector of the SVD unfolding in dependence of k.
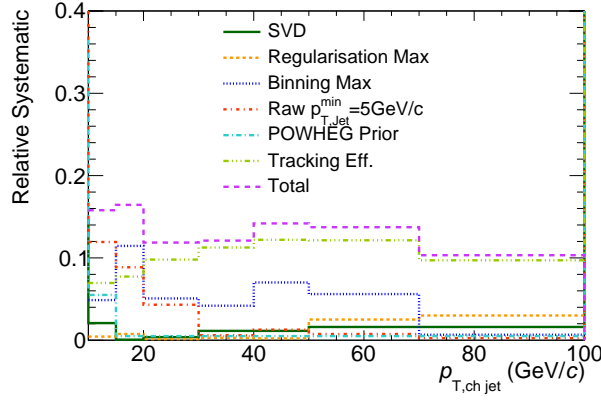
Figure 5.42. – The systematic uncertainties that originate from the different categories of unfolding variations as well as the tracking efficiency uncertainty. The combined uncertainty ("total") for the unfolding procedure and the tracking efficiency is also provided.

The regularisation and the prior uncertainty are given by Eqn. 5.5.10 and Eqn. 5.5.9 as for the Iterative Bayesian unfolding. The statistical uncertainty is, in this context, defined as the sum of the statistical uncertainties of all bins of the unfolded spectrum in dependence of k. In Fig. 5.41a, it can be seen that the total uncertainty is minimal for $k = 4$. Since the Pearson matrices show no significant correlations for $k = 3, \ldots, 5$ in the off-diagonal elements (see Fig. C.6), a regularisation parameter of $k = 4$ is chosen for the unfolding procedure. The unfolded b-jet spectra for $k = 4$ divided by respective distributions for different regularisation parameters are shown in Fig. 5.40b.

Fig. 5.42 provides the resulting systematic uncertainties for the unfolding procedure for every variation category. The largest contribution at medium and large $p_{T,ch\ jet}$ originates from the tracking efficiency uncertainty, which is at about 12 % at medium $p_{T,ch\ jet}$. The largest contribution at small $p_{T,ch\ jet}$ is the truncation uncertainty, which reaches 12 % for the lowest $p_{T,ch\ jet}$ bin. The total systematic uncertainty is slightly decreasing from about 16 % at low $p_{T,ch\ jet}$, 14 % at medium and 10 % at large $p_{T,ch\ jet}$.

5.6.5. Summary of systematic uncertainties

In this section, the results for the Barlow tests and the deviations that arise from the systematic variations discussed in the previous sections will be summarised. The different sources of systematic uncertainties on the performance-corrected spectrum will be classified as significant or insignificant based on the Barlow test defined in Eqn. 5.6.2. Only significant uncertainties are considered for the estimation of the final systematic uncertainty.

The following systematic variations with respect to uncertainties of the performance-corrected spectrum have been proven to be *insignificant* according to the Barlow test:

- the V^0-jet uncertainty (cat.1, see Sect. 5.6.2),

- the consideration of the MC template statistics (cat.1, see Sect. 5.6.3),

- the variation of the tagger working point (cat.1, see Sect. 5.6.3) and

- the variation of the track selection (cat.2, see Sect. 5.6.3).

The classification according to category 1 (cat.1) and category 2 (cat.2), as defined at the beginning of this section, are provided in brackets. The relative systematic deviations and the corresponding Barlow tests are summarised in Fig. 5.43a and Fig. 5.43c.
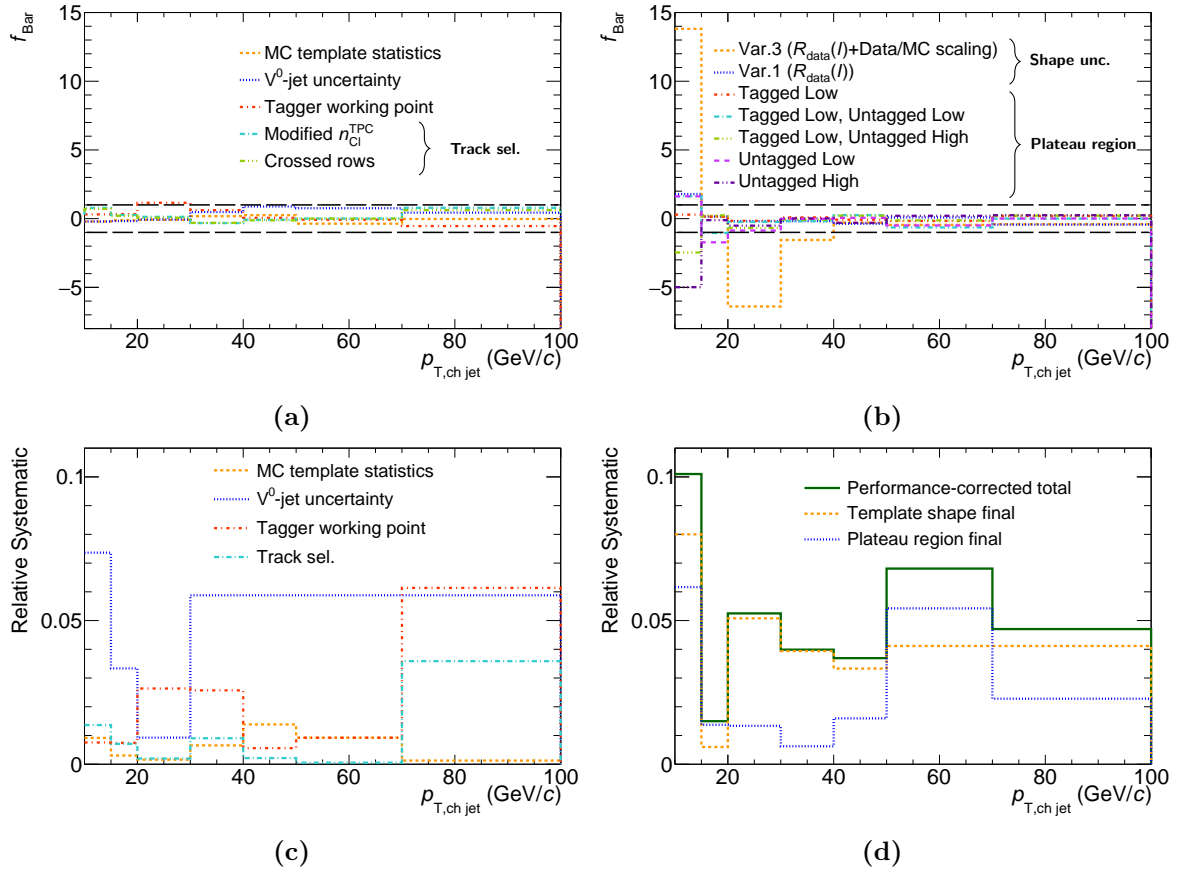
Figure 5.43. – The results for the systematic uncertainties on the performance corrected spectrum. **a)** The Barlow test for all variations that have negligible effect. **b)** The Barlow test for all variations with significant effect. **c)** The relative uncertainties for all variations that have negligible effect. **d)** The relative uncertainties for all variations with significant effect.

The uncertainties which have been identified as *significant* according to the Barlow test are

- the shape uncertainties (cat.1, see Sect. 5.6.1) and

- the variation of the plateau region (cat.1, see Sect. 5.6.3).

Only the latter uncertainties are considered for the final systematic uncertainties. The corresponding Barlow tests and relative uncertainties are summarised in Fig. 5.43b and Fig. 5.43d. It can be seen that both variations are especially significant at comparatively low $p_{\mathrm{T,ch\ jet}}$, where also the statistical uncertainties are smallest. At the same time, it has been shown in the dedicated sections above that variations with respect to the Sd_{xy} resolution as well as the effect of the fit-range averaging are particularly significant for $p_{\mathrm{T,ch\ jet}} < 30\,\mathrm{GeV}/c$. Consequently, the significance of both systematic uncertainties is not only driven by the size of the statistical uncertainties but has also physical reasons.

The significant uncertainties of cat.1 are combined to a single uncertainty $\Delta_{\mathrm{pc,\ cat1.}}$ via adding the individual contributions in quadrature. Since the unfolding procedure is sensitive to the steepness of the smeared data, the lower ($x_{\mathrm{low}} = x_{\mathrm{Def}} - \Delta_{\mathrm{pc,\ cat1.}}$) and the upper boundaries ($x_{\mathrm{up}} = x_{\mathrm{Def}} + \Delta_{\mathrm{pc,\ cat1.}}$) of the default b-jet spectrum with respect to the $\Delta_{\mathrm{pc,\ cat1.}}$ uncertainties are propagated separately through the unfolding procedure. This approach entails asymmetric systematic uncertainties for the final spectrum. Similar as for the default unfolding procedure, toy samples are generated from both boundaries. Thereby, the respective bin content is varied according to a Gaussian distribution centered at x_{low} (and x_{up} accordingly) with the standard deviation $\sigma_{\mathrm{Default}}$. All toys are unfolded and the mean of the unfolded spectra are adopted
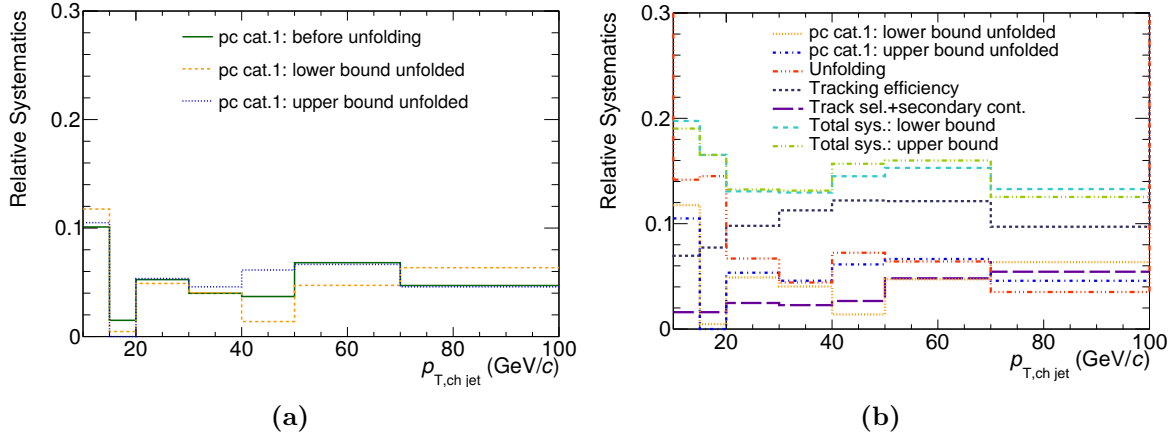
(a)

(b)

Figure 5.44. – **a)** The upper and the lower boundaries of the systematic uncertainties on the performance-corrected spectrum of cat.1 after the unfolding compared to the original cat.1 uncertainty. **b)** All significant contributions for the estimation of the final systematic uncertainties. See text for more details.

as the upper and the lower boundaries of the respective systematic uncertainties on the final b-jet spectrum. The unfolded $\Delta_{\text{pc, cat1.}}$ for the upper and the lower boundary are compared to the original $\Delta_{\text{pc, cat1.}}$ before the unfolding in Fig. 5.44a. The differences between the upper and the lower boundaries after the unfolding and between the latter two and the original cat.1 uncertainties are about 1 to 2 %.
The uncertainties $\Delta_{\text{pc, cat1.}}$ are combined with the unfolding uncertainties Δ_{unfold} according to Eqn. 5.6.3. As all cat.2 uncertainties have been identified as negligible, they are not considered for the final uncertainties (i.e. $\Delta_{\text{pc, cat.2}} = 0$).

The determination of two systematic variations was beyond the scope of this analysis: uncertainties with respect to the tracking resolution and with respect to the secondary vertex contamination are not addressed in this thesis. Thereby, the latter uncertainty corrects for an improper MC description of jet momentum smearing caused by secondary tracks. In [21] it has been found, that the respective uncertainties are small ($\lesssim 2.5\,\%$) at low $p_{\text{T,ch jet}}$ in comparison to other sources, whereas for $p_{\text{T,ch jet}} > 50\,\text{GeV}/c$, they reach about the same magnitude as the uncertainties on the performance-corrected spectrum in this analysis (see Fig. 5.44b). As a rough estimate for preliminary results of the b-jet spectrum in 13 TeV pp collisions, the uncertainties from [21] have been adopted for both sources. These have been added in quadrature to the uncertainties from the unfolding and on the performance-corrected spectrum to obtain the final systematic uncertainties.

As one can see in Fig. 5.44b, the final systematic uncertainties are dominated by the unfolding systematic uncertainties ($\sim 14\,\%$) at low $p_{\text{T,ch jet}}$ and by the tracking efficiency ($10 - 12\,\%$) at medium and large $p_{\text{T,ch jet}}$. The uncertainties on the performance-corrected spectrum contribute to the final systematic uncertainty with about $4 - 6\,\%$ for $p_{\text{T,ch jet}} > 20\,\text{GeV}/c$ and $10 - 12\,\%$ for $p_{\text{T,ch jet}} < 15\,\text{GeV}/c$.
The statistical and all systematic uncertainties on the final b-jet spectrum for low, medium and large jet momenta are listed in Tab.C.1.

5.7. Results and discussion

The final result for the p_T-differential cross section of charged-particle b jets in 13 TeV pp collisions, which has been fully corrected according to Eqn. 5.0.1, is shown in Fig. 5.45a. It is compared to a simulation with PYTHIA 8.307 using the Monash 2013 tune [148]. Further technical details on the settings of the PYTHIA simulation are listed in Appendix C.5.1. The PYTHIA simulation significantly overestimates the data at low $p_{T,\text{ch jet}}$ by up to a factor of 2.5. The discrepancy decreases with increasing $p_{T,\text{ch jet}}$ until PYTHIA and the data are compatible for $p_{T,\text{ch jet}} > 70\,\text{GeV}/c$.

A similar trend – an overestimation at low $p_{T,\text{ch jet}}$ with the simulation approaching the data at high $p_{T,\text{ch jet}}$ – has also been observed by CMS for inclusive b jets at 7 TeV and 13 TeV (see Sect. 4). Since different PYTHIA versions and tunes have been applied in the two analyses, a more quantitative comparison of the results would, however, not be sensible.

The overestimation of the b-jet cross section by PYTHIA which is observed in this thesis is more significant than for charged-particle jets of inclusive flavours for the same collisions system and energy (see Sect. 4). This indicates particular deficiencies of PYTHIA with respect to the description of heavy-flavour production.

The relative fraction of charged-particle beauty jets over jets of inclusive flavours is shown in Fig. 5.45b in comparison to the corresponding PYTHIA simulation. Thereby, the data for jets of inclusive flavours has been taken from the results presented in [22]. The systematic and statistical uncertainties of both measurements are assumed to be fully uncorrelated. The b-jet fraction increases from about 1 % at low $p_{T,\text{ch jet}}$ up to about 5 % at high $p_{T,\text{ch jet}}$. PYTHIA predicts a larger b-jet fraction for $p_{T,\text{ch jet}} \leq 30\,\text{GeV}/c$ but is compatible with the data at larger jet momenta.

As it was stated before, PYTHIA is a LO event generator with next-to-leading logarithmic accuracy. This is why, its description of NLO processes exhibits known deficiencies like the overprediction of the $g \rightarrow b\bar{b}$ splitting rate (see Sect. 2.2.6). Measurements indicate that the production of beauty quarks at LHC energies is dominated by NLO processes (see Sect. 2.2.2) which is why the discrepancies between data and simulation observed in this thesis meet the expectations.

For further analyses, it will be informative to test predictions from pQCD calculations by comparing the b-jet cross section to a simulation with POWHEG+PYTHIA 8. A respective comparison will necessitate a correction for the underlying event as it is performed e.g. in [21]. In Sect. 2.2.6, it was noted that the description of colour-coherence effects has an influence on predictions of the $g \rightarrow b\bar{b}$ splitting rate by PYTHIA. A comparison of the b-jet spectrum to predictions by the Vincia shower algorithm could provide additional insights with respect to the modeling of heavy-flavour production in this respect.

Fig. 5.46a compares the p_T-differential b-jet cross section for 13 TeV to the one for 5.02 TeV presented in [21]. While the spectrum at 5.02 TeV is corrected for the underlying event, this is not the case for the spectrum at 13 TeV. According to measurements for charged-particle jets of inclusive flavours in [16], this can lead to significant additional differences of larger than 10 % at low $p_{T,\text{ch jet}}$ while only slight deviations are to be expected at large $p_{T,\text{ch jet}}$. Going from low to large jet momenta, the b-jet cross section at 13 TeV is larger by a factor of about two to six than the respective spectrum at 5.02 TeV. The relative beauty-jet fractions, which are compared in Fig. 5.46b, are compatible with each other at both centre-of-mass energies.

The refinements of the method for the performance estimation of the b-jet tagger, that have been developed in the course of this thesis, have succeeded in stabilising the results for the b-jet cross section at low jet momentum. Because of this, the reach of the measurement could be extended from $20\,\text{GeV}/c$ to $10\,\text{GeV}/c$ in comparison to the IP method of the 5.02 TeV analysis. In comparison to the combined results of the IP and SV method of the latter analysis,
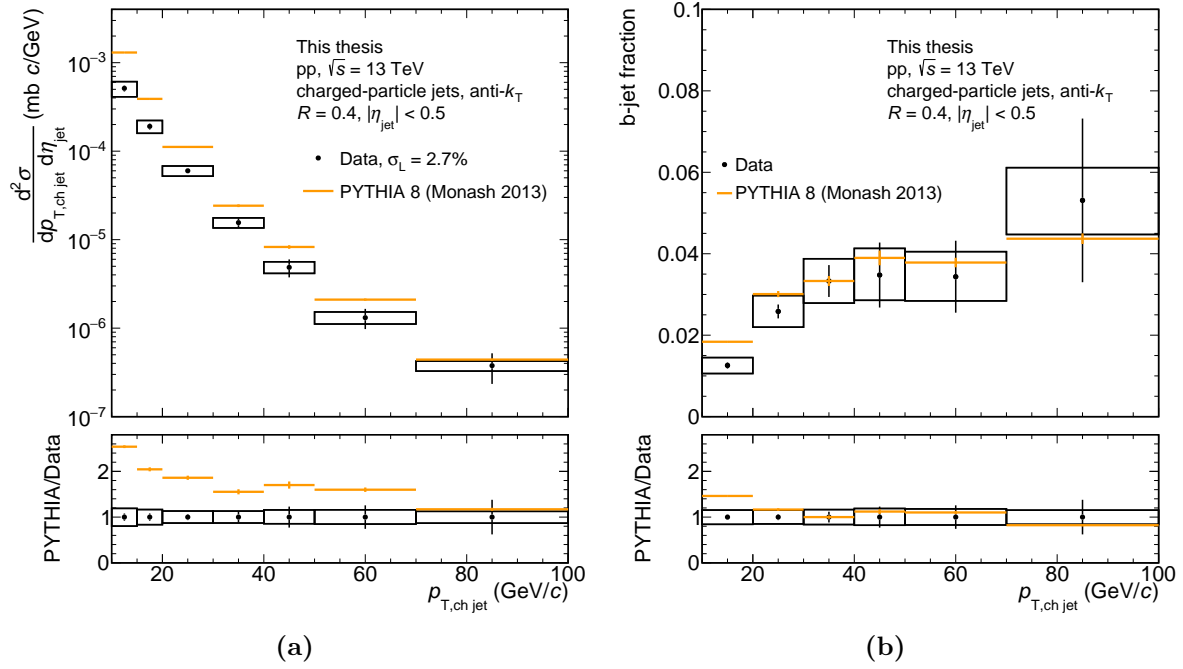
(a) **(b)**

Figure 5.45. – Comparison of spectra for charged-particle b jets in pp collisions with $\sqrt{s} = 13\,$TeV to a corresponding simulation with PYTHIA 8.3. Statistical uncertainties are shown as error bars whereas the systematic uncertainties of the data are provided as boxes. **a)** Comparison of the p_T-differential cross sections. The additional normalisation uncertainty of the data is $\sigma_L = 2.7\,\%$ [37]. **b)** Comparison of the relative fractions of charged-particle b jets over jets of inclusive flavours in dependence of $p_{T,\text{ch jet}}$. The cross section for jets of inclusive flavours from [22] are used as the baseline for the generation of the ratio.
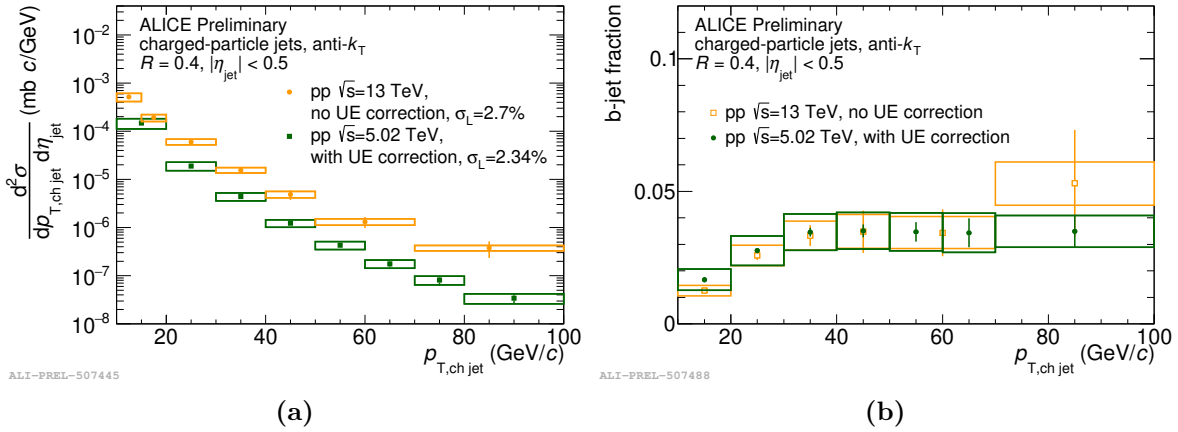

(a) **(b)**

Figure 5.46. – Comparison of spectra for charged-particle b jets in pp collisions with $\sqrt{s} = 13\,$TeV and $\sqrt{s} = 5.02\,$TeV. Statistical uncertainties are shown as error bars whereas systematic uncertainties are provided as boxes. The results for $\sqrt{s} = 5.02\,$TeV are taken from [21]. **a)** Comparison of the p_T-differential cross sections. The respective normalisation uncertainties σ_L [36, 37] of the data are quoted in the legend. **b)** Comparison of the relative fractions of charged-particle b jets over jets of inclusive flavours in dependence of $p_{T,\text{ch jet}}$. The cross section for jets of inclusive flavours from [22] are used as the baseline for the generation of the ratio for $\sqrt{s} = 13\,$TeV.

the improvements of the analysis strategy allowed for finner binning in jet momentum for $p_{\text{T,ch jet}} \leq 20\,\text{GeV}/c$ for about the same available statistics.

6. Exploring the Lund Plane for b jets

In Sect. 2.2.4, the ALICE measurement [15, 159] of the dead-cone effect for charm quarks via the analysis of D^0 jets has been introduced. It was noted, that the full reconstruction of D^0 mesons was crucial for the success of the applied analysis technique. This follows from the fact that, the decay products of heavy-flavour mesons fill the signal region of hard-collinear small-angle gluon branchings and thus conceal the dead-cone effect if they are not rejected [121].

In principle, it should be possible to uncover the dead-cone effect for beauty quarks by exploiting the Lund Plane for b jets as described in [82]. However, a full reconstruction of beauty hadrons is not possible with ALICE data for LHC Run 1 and 2 [98]. Whether a direct measurement of the dead-cone effect for b quarks is nevertheless possible with the proposed method and available ALICE data will be investigated in a simulation study with PYTHIA 8 in this chapter.

6.1. Analysis strategy

The beauty-jet Lund Plane has been simulated for pp collisions with a centre-of-mass energy of $\sqrt{s} = 13$ TeV using the event generator PYTHIA 8.240 with the Monash 2013 tune. For details on the simulation settings, it is referred to Appendix C.5.2. Charged-particle jets with a radius $R = 0.4$ and a jet transverse momentum of $30 < p_{\text{T,ch jet}} < 100$ GeV/c are identified with the anti-k_{T} algorithm and declustered using the Cambridge-Aachen algorithm to fill the Lund Plane as detailed in Sect. 2.2.5. The E-scheme recombination is applied and only jets within the ALICE acceptance are considered, which implies $|\eta_{\text{Jet}}| < 0.5$.

The simulation is designed to mimic an experimental measurement without the technical opportunity to identify beauty hadrons. This is why, beauty jets are defined as jets containing decay particles of $B^{\pm}$ and B^0 mesons. These particles can experimentally be identified as tracks with a large displacement from the primary vertex (see Sect. 4). Consequently, the definition that is used for this study differs from the one that is used for the reconstruction of beauty jets in Sect. 5.

In total, the Lund Plane is simulated independently for four different categories:

1. b jets with unstable B mesons, i.e. splittings within beauty jets *including* b-meson decay contributions are considered;

2. b jets with stable B mesons, i.e. splittings within beauty jets *excluding* B-meson decay contributions are considered. In this case, b jets are defined as jets containing a $B^{\pm}$ or B^0 meson;

3. B-meson decays, i.e. *only* splittings within beauty jets from B-meson decay contributions are considered. Technically, this means that only splittings are accepted for which both branches contain a B-meson decay product; and

4. inclusive jets, i.e. splittings within jets of inclusive flavours are considered.

Thereby, the Lund Plane for b jets with unstable B mesons shall substitute a realistic measurement based on ALICE data for LHC Run 2. Inclusive jets are to a large extend dominated by light-quark and gluon-initiated jets. Therefore, the simulation for inclusive jets is assumed

to measure the emission probability of light flavours and gluons. A comparison of the Lund Plane for inclusive jets to the Lund Plane for b-jets with stable B meson is expected to unveil the dead-cone effect with clear manifestation. Thereby, it needs to be taken into account, that also colour-charge effects induce differences between the radiation probabilities for quarks and gluons (see Sect. 2.2.4). Simulations by MC event generators discussed in [15, 159], however, show that in the absence of mass effects, quark jets fragment more collinearly than inclusive jets. Consequently, differences from colour-charge effects result in an *increased* radiation probability at low emission angles for b jets in comparison to inclusive jets. Thus, a *decrease* of the low-angle radiation probability for b jets in comparison to inclusive jets is a clear signature of the dead-cone effect.

The Lund Plane for B-meson decays is utilised to study the possibility to perform a multi-dimensional subtraction of the B-meson decay contributions from the b-jet Lund Plane for unstable B mesons to recover the Lund Plane for stable B mesons.

6.2. Simulation results

The results for the simulations of the four Lund-Plane categories are shown in Fig. 6.1. When comparing the Lund Planes for inclusive jets in Fig. 6.1b to the one for b jets with stable B mesons in Fig. 6.1d, it can be recognised that splittings in the latter are suppressed at $\ln(k_{\mathrm{T}}) \gtrsim -1 \,\mathrm{GeV}/c$ and $\ln(1/\theta) \gtrsim 1.5$ with respect to splittings in the former. However, there is no significant difference between the Lund Plane for inclusive jets and the Lund Plane for b jets with unstable B mesons in Fig. 6.1a, indicating that the signal region is smeared by decay contributions. The Lund Plane for B-meson decays in Fig. 6.1c shows a considerable abundance of splittings in the hard-collinear regime and thus supports this assumption.

Assuming that a method could be found in the future to identify B-meson decay splittings, it has been studied whether the Lund Plane for b jets with stable B mesons can be recovered by the subtraction of the Lund Plane for B-meson decays from the b-jet Lund Plane for unstable B mesons. The corresponding Lund Plane after the subtraction of decay contributions is shown in Fig. 6.1e. The multiplicity of splittings in the region of interest is decreased significantly.

For better visualisation of the dead-cone effect, the relative difference of the emission probabilities for b jets $P^{\mathrm{b}}(\ln(1/\theta))$ and inclusive jets $P^{\mathrm{inc}}(\ln(1/\theta))$ is determined as

$$Q_\theta = \frac{P^{\mathrm{b}}(\ln(1/\theta)) - P^{\mathrm{inc}}(\ln(1/\theta))}{P^{\mathrm{inc}}(\ln(1/\theta))} \tag{6.2.1}$$

in dependence of the angle θ and energy of the radiator E_{rad}, following the strategy in [82]. Thereby, the emission probabilities are obtained from projections of the respective Lund Planes onto the $\ln(1/\theta)$ axis. Non-perturbative effects from hadronisation processes are rejected by considering only splittings with a transverse momentum k_{T} above a certain value. The dead-cone effect is expected to manifest as $Q_\theta < 0$ for low-angle splittings.

The emission probabilities for radiator energies of $20 < E_{\mathrm{rad}} < 40 \,\mathrm{GeV}$ and $40 < E_{\mathrm{rad}} < 60 \,\mathrm{GeV}$ as well as a transverse momentum of $k_{\mathrm{T}} > 0.2 \,\mathrm{GeV}/c$ are shown in Fig. 6.2 for all splitting categories defined above. It can be seen, that low-angle splittings within inclusive jets are more frequent than within beauty jets with stable B mesons. As expected from the discussion in Sect. 2.2.4, this difference is more significant at smaller energies of the radiator which is why the interval $20 < E_{\mathrm{rad}} < 40 \,\mathrm{GeV}/c$ has been chosen for all further investigations.

The Q_θ distributions for the Lund Planes of the b-jet splitting categories as well as for the b-jet Lund Plane after the subtraction of decay contributions are presented in Fig. 6.3 for four selections on k_{T}. The variation of the k_{T} selection criterion merely limits the statistics at small emission angles and does not influence the individual shapes of the distributions. This demonstrates the
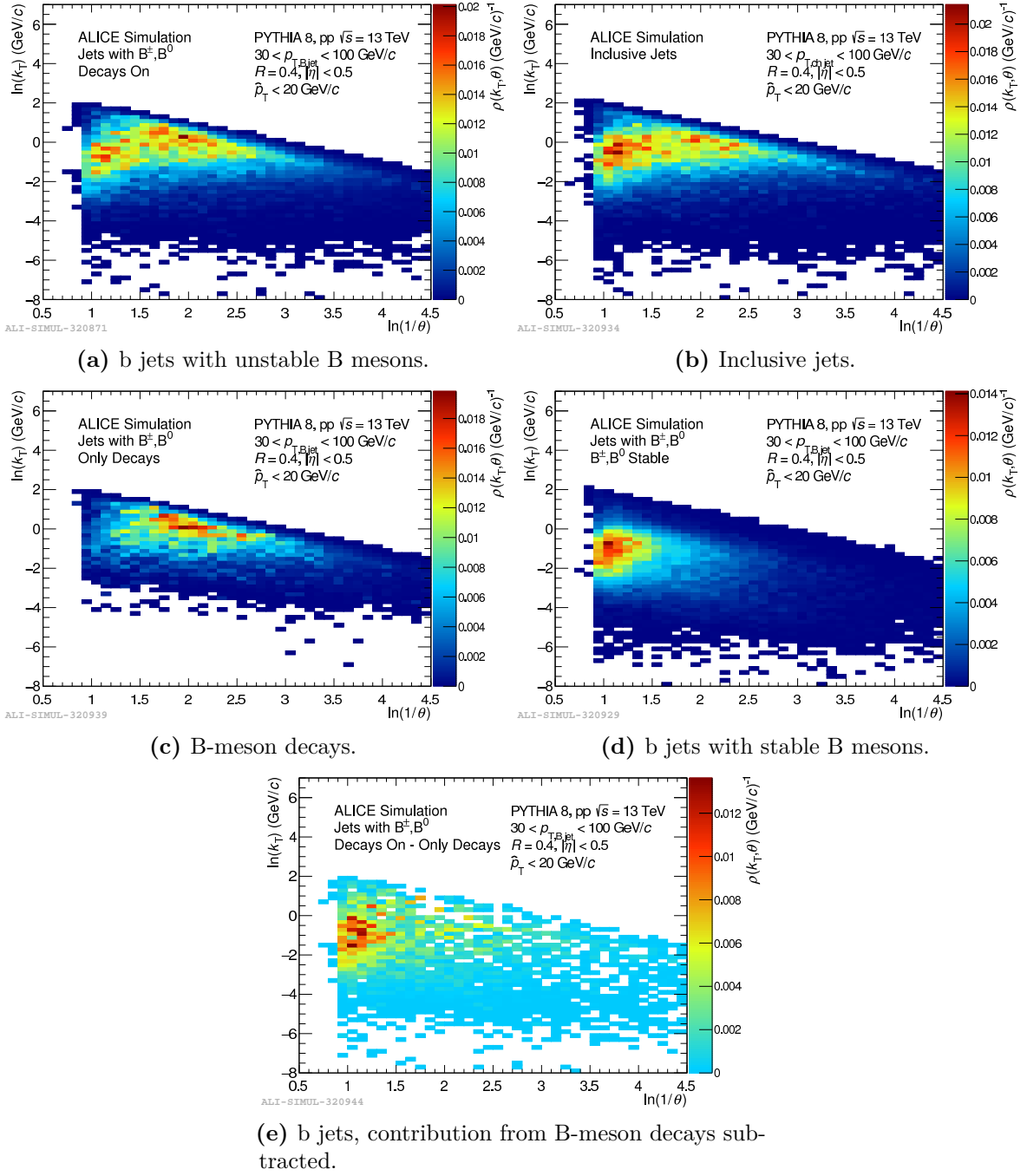
(a) b jets with unstable B mesons.

(b) Inclusive jets.

(c) B-meson decays.

(d) b jets with stable B mesons.

(e) b jets, contribution from B-meson decays subtracted.

Figure 6.1. – Simulations with PYTHIA 8 of the Lund Planes for jets with a transverse momentum of $30 < p_{\mathrm{T,ch\ jet}} < 100\,\mathrm{GeV}/c$ and different splitting categories that are detailed in the text.
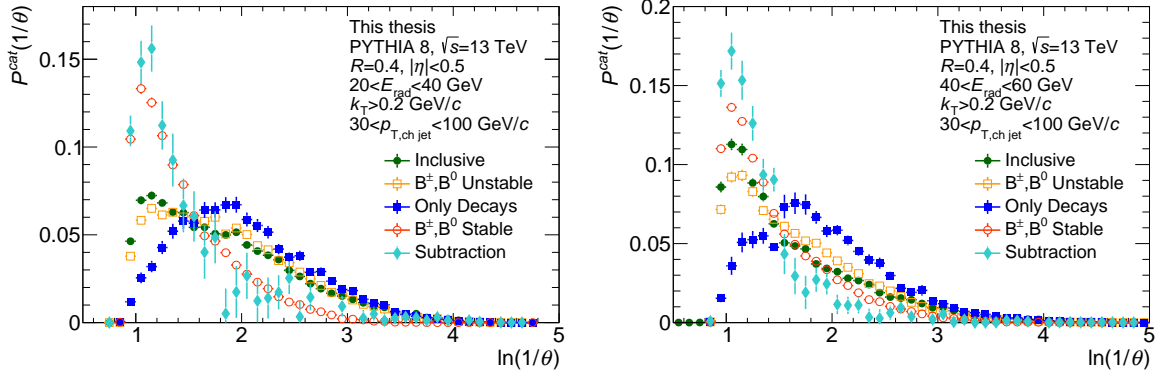
Figure 6.2. – Emission probability $P^{cat}(\ln(1/\theta))$ in dependence of the emission angle θ for all splitting categories *cat* defined in the text. The results are shown for a jet transverse momentum of $30 < p_{\mathrm{T,ch\ jet}} < 100\,\mathrm{GeV}/c$ and an energy of the radiator of $20 < E_{\mathrm{rad}} < 40\,\mathrm{GeV}$ (left) as well as $40 < E_{\mathrm{rad}} < 60\,\mathrm{GeV}$ (right).
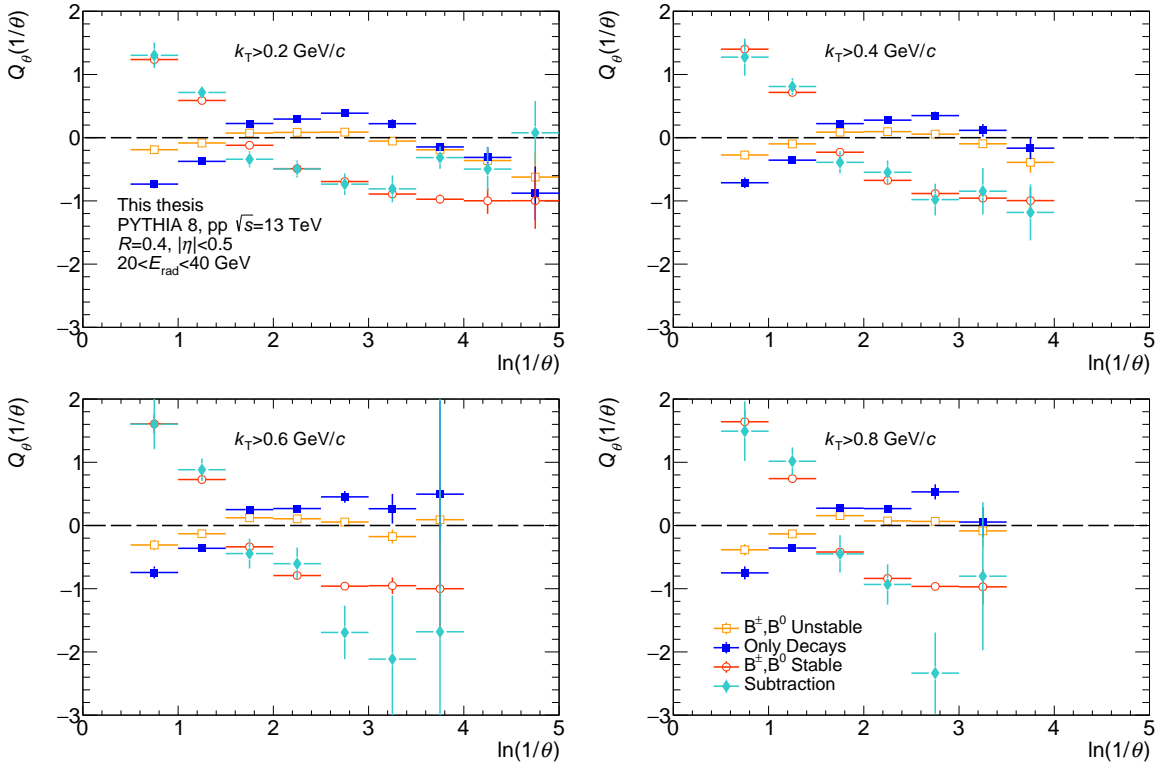

Figure 6.3. – Relative difference of the emission probabilities within beauty jets of different categories and inclusive jets. The results are shown for a jet transverse momentum of $30 < p_{\mathrm{T,ch\ jet}} < 100\,\mathrm{GeV}/c$, an energy of the radiator of $20 < E_{\mathrm{rad}} < 40\,\mathrm{GeV}$ and for different selections on the transverse momentum k_{T} of the splittings.
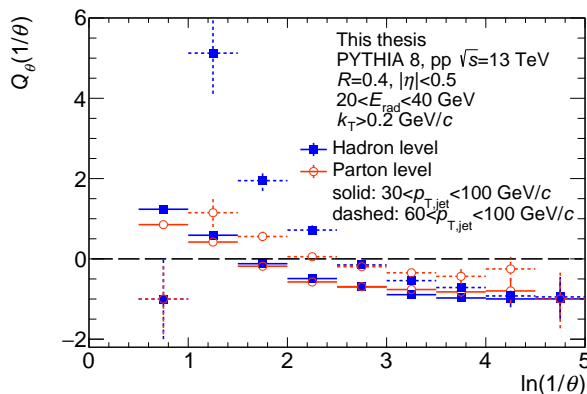
Figure 6.4. – Comparison of the relative difference between the emission probabilities within beauty jets with stable B mesons and within inclusive jets for $\ln(k_T) > 0.2$ at hadron level and at parton level. Solid lines represent the results for $30 < p_{T,\text{ch jet}} < 100 \, \text{GeV}/c$ while dashed lines are for $60 < p_{T,\text{ch jet}} < 100 \, \text{GeV}/c$.

robustness of the method with respect to hadronisation effects. The distributions for b jets with stable B mesons show a significant suppression below one for small angles of $\ln(1/\theta) \gtrsim 1.5$ – the dead-cone effect. On the other hand, the distributions for b jets with unstable B mesons are compatible with one. It is thus confirmed that B-meson decays conceal the dead-cone effect in the Lund Plane for measurement conditions that are realistic for ALICE during LHC Run 2. By the subtraction of the B-meson decay contributions from the b-jet Lund Plane for unstable B mesons, the Q_θ distribution for stable B mesons can be regained.

In the following, the sensitivity of the results with respect to hadronisation effects shall be further investigated. In this sense, the former *hadron-level* simulation, i.e. the accumulation of splittings after hadronisation, is complemented by a simulation of the Lund Plane at *parton-level* which is obtained by omitting the event evolution of PYTHIA after the parton shower [82]. The results for the corresponding Q_θ distributions are compared in Fig. 6.4. The predictions of the hadron-level and the particle-level simulations diverge for large jet momenta of $60 < p_{T,\text{ch jet}} < 100 \, \text{GeV}/c$. However, no considerable difference can be observed for the phase space of interest, i.e. $30 < p_{T,\text{ch jet}} < 100 \, \text{GeV}/c$, for the least strict k_T selection that has been considered and $\ln(1/\theta) \gtrsim 1.5$. Thus, it can be concluded that the suppression of the Q_θ distributions below one for b-jets with stable B mesons is a solid signature of the dead-cone effect.

6.3. Discussion

The simulation study that has been presented in this thesis confirms the general suitability of the method presented in [82] for investigations of the dead-cone effect. This conclusion is based on the observation of a distinct suppression of the relative difference of the emission probability for beauty and inclusive jets below one for beauty jets with stable B mesons. Within the statistical uncertainties, this suppression shows no difference between parton-level and hadron-level simulations in the investigated region of phase space. This underlines the robustness of the method with respect to the influence of non-perturbative hadronisation processes.

However, the results also show that, under realistic measurement conditions for the ALICE experiment in LHC Run 2, the signal region for the dead-cone effect is covered by B-meson decay splittings. For the simulation results, the Lund Plane for stable B mesons could be regained by subtracting the B-meson decay contributions. In case of a sensible measurement, it would nonetheless be necessary to identify the decay contributions in the data since a correction based solely on simulations would rely significantly on the correct description of the splitting kinematics

by the utilised event generator.

For future analyses, it would be interesting to investigate the possibility to reject B-meson decay splittings during the declustering process based on the track impact parameter. Hadrons that originate from gluon emissions off the original b quark should be dominated by light-flavour particles. Thus, their decay products are expected to generate tracks with a small distance to the primary vertex. In contrast, decay products from B mesons reveal themselves by large-impact parameter tracks. Similar to the identification of B-meson splittings in this simulation study, decay splittings could thus be singled out by rejecting splittings for which both branches contain large-impact parameter tracks. The practicability of this method could by tested in detector-level simulations, e.g. with a combination of PYTHIA and GEANT.

Independent from the challenging conditions for ALICE data from LHC Run 2, the possibility to fully reconstruct B mesons in data from LHC Run 3 will open the opportunity to investigate the dead-cone effect for beauty quarks [98].

7. Summary

In this thesis, the p_{T}-differential beauty-jet cross section for pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ is measured for charged-particle jets with a transverse momentum of $10 < p_{\mathrm{T,ch\ jet}} < 100\,\mathrm{GeV}/c$ and a radius of $R = 0.4$ using data from the ALICE experiment. In addition, the prospects for exposing the dead-cone effect for beauty quarks by a measurement of the Lund Plane for beauty jets without the possibility to reconstruct B mesons are addressed in a MC simulation study.

In the first part of this thesis, beauty jets are selected based on the signed transverse impact-parameter significance Sd_{xy} of the tracks within a jet. The selection is performed per event and the selection criteria depend on the transverse momentum of the jet.
Every $p_{\mathrm{T,ch\ jet}}$ bin of the raw tagged data is corrected for the purity and the efficiency of the b-jet tagger, which are estimated from the b-jet fractions in data samples before and after the b-jet tagging. The b-jet fractions are obtained from template fits to distributions of the jet probability $\ln(\mathrm{JP})$ in the data. The fit templates for light-flavour, charm and beauty jets are simulated with a combination of the LO event generator PYTHIA 8.2 and GEANT 3. The performance corrected b-jet spectrum is unfolded for detector effects using Iterative Bayesian Unfolding.
To make the results for the performance estimation robust with respect to statistical fluctuations of the data as well as inaccuracies of the MC description of the fit templates, the method as it has been applied in former analyses [21] is optimised in two aspects. First, the fit results for every $p_{\mathrm{T,ch\ jet}}$ bin are averaged over the results for a selection of different fit ranges. In addition, the fit results for every fit range are averaged over an entity of individual fits for which the data is varied within its statistical uncertainties. The b-jet tagger is optimised to achieve an efficiency of about $60\,\%$ at a purity of about $40\,\%$, which is comparable to a former ALICE analysis [21], with a flat ratio of the tagging purity and the efficiency throughout the $p_{\mathrm{T,ch\ jet}}$ range.

Dedicated methods are developed to address the systematic effects resulting from an imprecise MC description of the $\ln(\mathrm{JP})$ fit templates as well as the contamination of the b-jet sample by jets containing V^0-decay particles. As a first step, a new method is introduced to consider discrepancies that arise from an inaccurate description of the Sd_{xy} distributions by the MC simulations. In this sense, an algorithm is developed that generates random Sd_{xy} values based on probability distributions for track properties. Jet objects are simulated as entities of tracks with corresponding Sd_{xy} values and the $\ln(\mathrm{JP})$ is calculated for every jet object. By manipulating the input histograms, the effect of modified Sd_{xy} distributions on the respective $\ln(\mathrm{JP})$ distributions can be investigated. The modified $\ln(\mathrm{JP})$ distributions are used for the sophisticated estimation of the template-shape uncertainties as well as to explore the influence of the modeling of the Sd_{xy} distributions on the stability of the $\ln(\mathrm{JP})$ template fits. It is confirmed, that an inaccurate description of the positive tails of the Sd_{xy} distributions for the individual jet flavours can lead to template-fit results that show a strong fit-range dependence.
For the first time in connection with a b-jet tagger, the contamination from V^0-decay particles is estimated from the real data by a dedicated V^0-jet tagger. A V^0-jet template is obtained, which is then fed to the procedure for the performance estimation as a fourth fit template. It is demonstrated that the $\ln(\mathrm{JP})$ distributions of V^0 jets are shifted towards larger $-\ln(\mathrm{JP})$ values than those for light-flavour jets and that V^0 jets contribute significantly to the $\ln(\mathrm{JP})$ distributions of b-tagged jets at low $-\ln(\mathrm{JP})$.

While the systematic uncertainty for the modeling of the Sd_{xy} distributions contributes significantly with up to $8\,\%$ at low $p_{\mathrm{T,ch\ jet}}$, the uncertainty from the V^0-jet contamination was found to be negligible with respect to the statistical uncertainties.

The results for the beauty-jet cross section are compared to a simulation with PYTHIA 8.3 which significantly overestimates the data at low $p_{\mathrm{T,ch\ jet}}$ by up to a factor of 2.5. The discrepancy between data and simulation is decreasing with increasing $p_{\mathrm{T,ch\ jet}}$. The PYTHIA prediction for the relative beauty-jet fraction is compatible with the data for the major part of the $p_{\mathrm{T,ch\ jet}}$ range. Considering that former measurements of the b-jet cross section are well described by calculations of the matrix-element generator POWHEG in combination with PYTHIA [21, 77], the discrepancy between the PYTHIA simulation and the data found in this thesis underlines the necessity for considering NLO processes for the accurate description of the absolute beauty-jet production.
The b-jet measurement is also compared to the results of a former ALICE analysis for pp collisions at $5.02\,\mathrm{TeV}$ [21]. The $p_{\mathrm{T,ch\ jet}}$-dependent b-jet fractions are found to be compatible for both measurements. Applying the refinements for the method of the performance correction mentioned above, the systematic uncertainties at low $p_{\mathrm{T,ch\ jet}}$ can be decreased significantly in comparison to the $5.02\,\mathrm{TeV}$ analysis. Consequently, the measurement presented in this thesis can be extended down to a jet momentum of $10\,\mathrm{GeV}/c$ in comparison to the lower boundary of $20\,\mathrm{GeV}/c$ for the part of $5.02\,\mathrm{TeV}$ analysis that applies similar methods as used in this thesis.
In this analysis, only one third of the available ALICE data for pp collisions at $13\,\mathrm{TeV}$ are analysed. The consideration of the full data sample will reduce the statistical uncertainties at large jet momenta while at the same time the estimation of systematic uncertainties that have been identified as negligible for the limited statistics in this thesis might gain in significance.
The results for the beauty-jet cross section extend an available measurement by the CMS collaboration from $74\,\mathrm{GeV}/c$ down to a transverse momentum of $10\,\mathrm{GeV}/c$. With this, the measurement discussed in this thesis leads the path to a further confrontation of predictions by pQCD calculations and MC event generators with data for b jets at low jet momenta. In this context, comparisons to predictions by POWHEG+PYTHIA 8 simulations as well as predictions by PYTHIA 8.3 for the Vincia shower algorithm could provide interesting insights.

In the second part of this thesis, the Lund Plane for b jets and jets of inclusive flavours is simulated with PYTHIA 8.3 for pp collisions at $13\,\mathrm{TeV}$. The relative difference of the gluon-emission probabilities within b jets and jets of inclusive flavours is investigated based on projections of the Lund Plane as recommended in [82]. While for b jets with hypothetical stable B mesons, the dead-cone effect manifests as a clear suppression of the low-angle radiation probability with respect to jets of inclusive flavours, no such effect is observed for b jets with unstable B mesons. Thus, it is proven that B-meson decay splittings fill the signal region of the dead-cone effect. However, it is also confirmed that the observation of the dead-cone effect for beauty quarks is feasible by a measurement of the b-jet Lund Plane if the beauty-hadron decay contributions can be rejected. In this context, a method is proposed which identifies B-meson decay contributions based on the transverse impact parameter of the decay products and rejects them during the declustering process.
The above findings provide valuable information with respect to prospects of corresponding analyses for ALICE data from LHC Run 2 and Run 3. Only for data from Run 3, the identification of B-meson decay splittings by the full reconstruction of B mesons will be possible [98].

To conclude, former methods for the b-jet tagging are optimised to measure the b-jet cross section down to a $p_{\mathrm{T,ch\ jet}}$ of $10\,\mathrm{GeV}/c$. In the context of former analyses, a comparison of the results to PYTHIA 8 simulations confirms that the consideration of NLO processes is essential for a sensible description of the absolute b-jet production. Furthermore, it is proven in a PYTHIA 8 simulation study that the dead-cone effect for b quarks can be exposed by the measurement of the b-jet Lund Plane, provided that decay contributions from B mesons are identified and rejected.

8. Zusammenfassung

In dieser Arbeit wurde der Wirkungsquerschnitt für die Produktion von Beauty-Jets in Proton-Proton-Kollisionen bei einer Schwerpunktsenergie von 13 TeV gemessen. Jets mit einem Transversalimpuls von $10 < p_{\mathrm{T,ch\ jet}} < 100\,\mathrm{GeV}/c$ und einem Radius von $R = 0.4$ wurden über geladener Teilchen aus Daten des ALICE Experiments rekonstruiert. Darüber hinaus wurde mithilfe von Monte-Carlo-Simulationen überprüft, inwieweit der sogenannte Dead-Cone-Effect für Beauty-Quarks über die Messung der Lund-Plane für Beauty-Jets sichtbar gemacht werden kann, obwohl die zugrunde liegenden Daten keine vollständige Rekonstruktion von B-Mesonen erlauben.

In dem ersten Teil dieser Arbeit wurden Beauty-Jets über die Distanz der im Jet enthaltenen Teilchenspuren zum primären Kollisionspunkt - gemessen durch den Stoßparameter Sd_{xy} – identifiziert. Die Selektion erfolgte ereignisbasiert und in Abhängigkeit von dem Transversalimpuls des jeweiligen Jets. Die einzelnen $p_{\mathrm{T,ch\ jet}}$-Intervalle des b-Jet-Spektrums wurden entsprechend der von dem b-Jet-Selektionsalgorithmus erzielten Reinheit und Effizienz korrigiert. Die Effizienz und die Reinheit wurden über die relativen Anteile der b-Jets in dem Datensatz vor und nach der b-Jet-Selektion bestimmt. Die Anteile der b-Jets folgten dabei aus den Ergebnissen von Template-Fits mit den Verteilungen der sogenannten Jet-Wahrscheinlichkeit ln(JP) für verschiedene Jet-Flavour an die zugehörigen Daten für Jets aller Flavour. Die Verteilungen für die Jets mit unterschiedlichem Flavour wurde mit PYTHIA 8 und GEANT 3 simuliert. Das effizienzkorrigierte b-Jet-Spektrum wurde mit der Methode des Iterative Bayesian Unfolding entfaltet, um Detektoreffekte zu kompensieren.

Im Vergleich zu der in vorherigen Analysen angewendeten Methode [21] wurden insbesondere zwei Arbeitsschritte optimiert, um die Ergebnisse robuster im Hinblick auf statistische Schwankungen der Daten sowie Ungenauigkeiten bei der Beschreibung der ln(JP)-Templates durch die Monte-Carlo-Simulationen zu machen. Zunächst wurden die Ergebnisse für die b-Jet-Anteile jedes $p_{\mathrm{T,ch\ jet}}$-Intervalls über die Ergebnisse von Fits über unterschiedliche Fit-Bereiche gemittelt. Darüber hinaus wurden die Daten für jeden einzelnen Fit innerhalb ihrer statistischen Unsicherheiten variiert und der Mittelwert der Ergebnisse von Fits an die variierten Daten bestimmt. Der b-Jet-Selektionsalgorithmus erzielte eine Effizienz von ca. 60% bei einer Reinheit von ca. 40%, was mit einer früheren ALICE-Analysen [21] vergleichbar ist. Das Verhältnis zwischen Reinheit und Effizienz ist über den Messbereich hinweg in etwa konstant.

Zur Bestimmung systematischer Effekte, welche aus einer ungenauen Beschreibung der ln(JP)-Fit-Templates durch die MC Simulationen resultieren, wurden spezielle Methoden entwickelt. Zunächst wurde eine neue Methode eingeführt, mithilfe derer Unstimmigkeiten untersucht werden können, die von einer fehlerbehafteten Modellierung der Sd_{xy}-Verteilungen durch die MC Simulationen resultieren. Es wurde ein Algorithmus entwickelt, der zufällige Sd_{xy}-Werte entsprechend vorhandener Wahrscheinlichkeitsverteilungen für Eigenschaften von Teilchenspuren generiert und auf diese Weise Jet-Objekte als Einheiten von Teilchenspuren simuliert. Die Verteilungen, welche als Basis für den Algorithmus dienen, können beliebig modifiziert und der Einfluss dieser Modifikationen auf die zugehörigen ln(JP)-Verteilungen untersucht werden. Die modifizierten ln(JP)-Verteilungen wurden verwendet, um die systematischen Unsicherheiten bezüglich der Form der Fit-Templates zu bestimmen, aber auch, um den Einfluss der Beschreibung der Sd_{xy}-Verteilungen auf die Stabilität der ln(JP)-Template-Fits zu beleuchten. Dadurch konnte gezeigt werden, dass eine fehlerhafte Beschreibung der Sd_{xy}-Verteilungen für $Sd_{xy} > 0$ für die Templates

der unterschiedlichen Jet-Flavour zu einer signifikanten Abhängigkeit der Fit-Ergebnisse vom Fit-Bereich führen kann.

Zum ersten Mal in Verbindung mit einem b-Jet-Tagger wurde die Kontamination der selektierten Jets durch V^0-Zerfallsteilchen auf der Basis von realen Daten mithilfe eines dedizierten V^0-Jet-Selektionsalgorithmus bestimmt. Zu diesem Zweck wurde ein V^0-Jet-Template erzeugt. Die verwendeten Selektionsbedingungen beruhten dabei auf jenen, die in [18, 149] angewandt wurden. Die V^0-Jet-Templates fungierten als viertes Fit-Template bei der Effizienzbestimmung.

Während die systematischen Unsicherheiten für die Beschreibung der Sd_{xy}-Verteilungen mit bis zu 8 % einen signifikanten Beitrag zur Summe der systematischen Unsicherheiten lieferten, zeigten sich die Unsicherheiten resultierend aus der V^0-Jet-Kontamination als vernachlässigbar im Vergleich zu den statistischen Unsicherheiten.

Die Ergebnisse für den Wirkungsquerschnitt der b-Jets wurden mit PYTHIA-Simulationen verglichen, welche die Daten bei niedrigen $p_{T,ch\,jet}$ mit bis zu einem Faktor von 2,5 deutlich überschätzen. Die Diskrepanz von Daten und Simulationen nimmt dabei mit zunehmendem $p_{T,ch\,jet}$ ab. Die PYTHIA-Simulationen für den relativen Anteil an b-Jets ist jedoch über einen großen Bereich der Messung mit den Daten kompatibel. Im Kontext früherer Messungen, bei denen eine gute Übereinstimmung des b-Jet-Wirkungsquerschnitts mit Berechnungen des Matrix-Element-Generators POWHEG in Kombination mit PYTHIA gefunden wurde [21, 77], unterstreicht die gefundene Diskrepanz zwischen PYTHIA-Simulationen und Daten, dass Prozesse höherer Ordnung zur Beschreibung der absoluten Produktion von b-Jets berücksichtigt werden müssen. Die Messung der b-Jets wurde ebenfalls mit den Ergebnissen einer Analyse [21] von ALICE für Proton-Proton-Kollisionen bei einer Schwerpunktsenergie von 5.02 TeV verglichen. Der $p_{T,ch\,jet}$-abhängige relative Anteil an b-Jets ist kompatibel für beide Messungen. Es zeigt sich, dass durch die oben beschriebene Weiterentwicklung der Methode für die Effizienzbestimmung die systematischen Unsicherheiten bei niedrigem $p_{T,ch\,jet}$ im Vergleich zu der 5.02 TeV-Analyse deutlich verringert werden konnten.

In dieser Analyse wurde nur etwa ein Drittel der verfügbaren Daten des ALICE-Experiments für Proton-Proton-Kollisionen bei 13 TeV analysiert. Eine Ausdehnung der Analyse auf den vollständigen Datensatz würde die statistischen Unsicherheiten bei hohen Transversalimpulsen deutlich verringern. Gleichzeitig würden systematische Unsicherheiten, welche in dieser Arbeit als vernachlässigbar identifiziert wurden, möglicherweise signifikant.

Die Messung für den Wirkungsquerschnitt der b-Jets erweitert eine frühere Messung der CMS Kollaboration hin zu niedrigen $p_{T,ch\,jet}$ von einem Transversalimpuls von 74 GeV/c auf einen Wert von 10 GeV/c. Die in dieser Arbeit präsentierte Analyse eröffnet damit die Möglichkeit weiterer Gegenüberstellungen von theoretischen Vorhersagen und b-Jet-Messungen bei niedrigen $p_{T,ch\,jet}$. In diesem Zusammenhang können Vergleiche mit den Vorhersagen von Simulationen mit POWHEG und PYTHIA 8 sowie Vergleiche mit PYTHIA 8.3 unter Verwendung des Vincia-Algorithmus für die Beschreibung des Partonschauers wichtige Erkenntnisse liefern.

In dem zweiten Teil dieser Arbeit wurde die Lund-Plane für b-Jets sowie Jets aller Flavour mit PYTHIA 8.3 für Proton-Proton-Simulationen bei einer Schwerpunktsenergie von 13 TeV simuliert. Die relative Differenz der Wahrscheinlichkeiten für die Gluonabstrahlung von b-Quarks in b-Jets und in Jets aller Flavour wurde untersucht. Dazu wurden, den Empfehlungen in [82] folgend, Projektionen der Lund-Plane auf die Winkelachse betrachtet. Es wurde gezeigt, dass die Abstrahlungswahrscheinlichkeit für b-Quarks in b-Jets unter der Annahme stabiler B-Mesonen im Vergleich zu Jets aller Flavour bei kleinen Winkeln unterdrückt ist. Dies kann als Manifestation des Dead-Cone-Effekts betrachtet werden. Für b-Jets mit instabilen B-Mesonen war dieses Verhalten nicht sichtbar. Folglich wurde bestätigt, dass der als Gluon-Abstrahlung fehlinterpretierte Zerfall der B-Mesonen diesen Effekt auf der Lund-Plane überdeckt. Nichtsdestotrotz wurde aufgezeigt, dass der Dead-Cone-Effekt für Beauty-Quarks auf der Lund-Plane für b-Jets nachgewiesen werden kann, wenn die Beiträge von B-Meson-Zerfällen herausgefiltert werden. In diesem Zusammenhang wurde eine Methode vorgeschlagen, mit der die Beiträge

von B-Meson-Zerfällen über die transversalen Stoßparameter der Zerfallsteilchen während des Entflechtungsprozesses der Gluon-Abstrahlungen unterdrückt werden können. Darüber hinaus wurde bestätigt, dass die verwendeten Methoden unempfindlich gegenüber dem Einfluss von nicht-perturbativen hadronischen Prozessen sind.

Die oben genannten Folgerungen liefern weitere wertvolle Informationen für potentielle zukünftige Analysen, welche Daten des ALICE-Experiments von LHC Run 2 und Run 3 verwenden. Lediglich für Daten von Run 3 können B-Meson-Zerfallsteilchen über die Rekonstruktion von B-Mesonen identifiziert werden [98].

Zusammenfassend wurde der Wirkungsquerschnitt für b-Jets mit niedrigem Transversalimpuls bis zu einer unteren Grenze von $10\,\text{GeV}/c$ gemessen. Anhand eines Vergleichs von PYTHIA-Simulationen mit dem gemessenen Wirkungsquerschnitt im Hinblick auf die Ergebnisse früherer Analysen wurde bestätigt, dass Prozesse höherer Ordnung für die akkurate Beschreibung der absoluten Produktion von Beauty-Jets berücksichtigt werden müssen. Darüber hinaus, wurde mithilfe von PYTHIA-Simulationen nachgewiesen, dass der Dead-Cone-Effect für Beauty-Quarks über die Messung der Lund-Plane für b-Jets enthüllt werden kann, solange eine Identifizierung und Unterdrückung der Beiträge von B-Meson-Zerfallsprodukten erfolgt.

Bibliography

[1] M. Aaboud et al., "Measurement of the nuclear modification factor for inclusive jets in Pb+Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV with the ATLAS detector", Phys. Lett. B **790**, 108–128 (2019).

[2] G. Aad et al., "The ATLAS Experiment at the CERN Large Hadron Collider", JINST **3**, Also published by CERN Geneva in 2010, S08003. 437 p (2008).

[3] G. Aad et al., "Measurement of the inclusive and dijet cross-sections of b^- jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector", Eur. Phys. J. C **71**, 1846 (2011).

[4] G. Aad et al., "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", Phys. Lett. B **716**, 1–29 (2012).

[5] G. Aad et al., "Performance of b-Jet Identification in the ATLAS Experiment", JINST **11**, P04008 (2016).

[6] R. Aaij et al., "Measurement of forward J/ψ production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV", JHEP **10**, [Erratum: JHEP 05, 063 (2017)], 172 (2015).

[7] R. Aaij et al., "Measurement of b hadron fractions in 13 TeV pp collisions", Phys. Rev. D **100**, 031102 (2019).

[8] K. Aamodt et al., "The ALICE experiment at the CERN LHC. A Large Ion Collider Experiment", JINST **3**, Also published by CERN Geneva in 2010, S08002. 259 p (2008).

[9] K Aamodt et al., "Alignment of the ALICE Inner Tracking System with cosmic-ray tracks", JINST **5**, P03003 (2010).

[10] J. Abdallah et al., "b tagging in DELPHI at LEP", Eur. Phys. J. C **32**, 185–208 (2004).

[11] B. Abelev et al., "Technical Design Report for the Upgrade of the ALICE Inner Tracking System", J. Phys. G **41**, 087002 (2014).

[12] B. Abelev et al., "Measurement of electrons from beauty hadron decays in pp collisions at $\sqrt{s} = 7$ TeV", Phys. Lett. B **721**, [Erratum: Phys.Lett.B 763, 507–509 (2016)], 13–23 (2013).

[13] B. B. Abelev et al., "Performance of the ALICE Experiment at the CERN LHC", Int. J. Mod. Phys. A **29**, 1430044 (2014).

[14] P. Abreu et al., "Hadronization properties of b quarks compared to light quarks in $e^+ e^- \to \overline{qq}$ from $183\,\mathrm{GeV}$ to $200\,\mathrm{GeV}$", Phys. Lett. B **479**, [Erratum: Phys.Lett.B 492, 398–398 (2000)], 118–128 (2000).

[15] S. Acharya et al., "Direct observation of the dead-cone effect in quantum chromodynamics", Nature **605**, [Erratum: Nature 607, E22 (2022)], 440–446 (2022).

[16] S. Acharya et al., "Measurement of charged jet cross section in pp collisions at $\sqrt{s} = 5.02\,\mathrm{TeV}$", Phys. Rev. D **100**, 092004 (2019).

[17] S. Acharya et al., "Measurement of the production of charm jets tagged with D^0 mesons in pp collisions at $\sqrt{s} = 7$ TeV", JHEP **08**, 133 (2019).

[18] S. Acharya et al., "K_S^0- and (anti-)Λ-hadron correlations in pp collisions at $\sqrt{s} = 13$ TeV", Eur. Phys. J. C **81**, 945 (2021).

[19]S. 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).

[20]S. Acharya et al., "Production of light-flavor hadrons in pp collisions at $\sqrt{s} = 7$ and $\sqrt{s} = 13$ TeV", Eur. Phys. J. C **81**, 256 (2021).

[21]S. Acharya et al., "Measurement of inclusive charged-particle b-jet production in pp and p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV", JHEP **01**, 178 (2022).

[22]S. Acharya et al., "Multiplicity dependence of charged-particle jet production in pp collisions at $\sqrt{\mathbf{s}} = \mathbf{13}$ **TeV**", Eur. Phys. J. C **82**, 514 (2022).

[23]S. Acharya et al., "Prompt and non-prompt J/ψ production cross sections at midrapidity in proton-proton collisions at $\sqrt{s} = 5.02$ and 13 TeV", JHEP **03**, 190 (2022).

[24]Acharya S. et al., "Anisotropic flow of identified particles in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$TeV", Journal of High Energy Physics **2018**, 6 (2018).

[25]Acharya S. et al., "Prompt D0, D+, and D*+ production in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV", Journal of High Energy Physics **2022**, 174 (2022).

[26]L. Adamczyk et al., "Energy Dependence of Moments of Net-proton Multiplicity Distributions at RHIC", Phys. Rev. Lett. **112**, 032302 (2014).

[27]J. Adolfsson et al., "The upgrade of the ALICE TPC with GEMs and continuous readout", JINST **16**, P03022 (2021).

[28]T. Adye, "Unfolding algorithms and tests using RooUnfold", in PHYSTAT 2011 (2011), pp. 313–318.

[29]C. Aidala et al., "Correlations of $\mu\mu$, $e\mu$, and ee pairs in $p+p$ collisions at $\sqrt{s} = 200$ GeV and implications for $c\bar{c}$ and $b\bar{b}$ production mechanisms", (May 2018) `https://arxiv.org/pdf/1805.04075.pdf`.

[30]G. Alexander et al., "A Study of b quark fragmentation into B0 and B+ mesons at LEP", Phys. Lett. B **364**, 93–106 (1995).

[31]A. Ali, "B decays – Introduction and Overview", in *B decays*, edited by S. Stone (World Scientific, Singapore et al., 1994), pp. 1–79.

[32]ALICE Collaboration, "AliPhysics framework", `https://github.com/alisw/AliPhysics`.

[33]ALICE Collaboration, "AliROOT framework", `https://github.com/alisw/AliRoot`.

[34]ALICE Collaboration, "Introduction to the Jet Framework", `http://alidoc.cern.ch/AliPhysics/master/READMEjetfw.html`, accessed on 19.09.22.

[35]ALICE Collaboration, "Preliminary Figure: ALI-PREL-341675", accessed on 21.09.2022, `https://alice-figure.web.cern.ch/node/17138`.

[36]ALICE Collaboration, "ALICE luminosity determination for pp collisions at $\sqrt{s} = 5$ TeV", ALICE Public Note, 2016.

[37]ALICE Collaboration, "ALICE 2016-2017-2018 luminosity determination for pp collisions at $\sqrt{s} = 13$ TeV", ALICE Public Note, 2021.

[38]"ALICE Time-Of-Flight system (TOF): Technical Design Report", tech. rep. (ALICE Collaboration, Geneva, 2000).

[39]S. Alioli et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", JHEP **06**, 043 (2010).

[40]S. Alioli et al., "Jet pair production in POWHEG", JHEP **04**, 081 (2011).

[41]J. Allen et al., "ALICE DCal: An Addendum to the EMCal Technical Design Report Di-Jet and Hadron-Jet correlation measurements in ALICE", tech. rep. (ALICE Collaboration, 2010).

[42] J. Alme et al., "The ALICE TPC, a large 3-dimensional tracking device with fast readout for ultra-high multiplicity events", Nucl.Instrum.Meth. **A622**, 316–367 (2010).

[43] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", Nucl. Phys. B **126**, 298–318 (1977).

[44] A. A. Alves et al., "The LHCb Detector at the LHC", JINST **3**, Also published by CERN Geneva in 2010, S08005 (2008).

[45] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", JHEP **07**, 079 (2014).

[46] J. Alwall et al., "MadGraph 5 : Going Beyond", JHEP **06**, 128 (2011).

[47] Y. S. Amhis et al., "Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018", Eur. Phys. J. C **81**, 226 (2021).

[48] R. Atkin, "Review of jet reconstruction algorithms", J. Phys. Conf. Ser. **645**, edited by A. S. Cornell and B. Mellado, 012008 (2015).

[49] M. Bahr et al., "Herwig++ Physics and Manual", Eur. Phys. J. C **58**, 639–707 (2008).

[50] R. D. Ball et al., "Parton distributions with QED corrections", Nucl. Phys. B **877**, 290–320 (2013).

[51] R. D. Ball et al., "Parton distributions for the LHC Run II", JHEP **04**, 040 (2015).

[52] R. Barlow, "Extended Maximum Likelihood", Nuclear Instruments and Methods in Physics Research **A297**, 496–506.

[53] R. Barlow, "Systematic errors: Facts and fictions", in Conference on Advanced Statistical Techniques in Particle Physics (July 2002), pp. 134–144.

[54] A. Beraudo et al., "Extraction of Heavy-Flavor Transport Coefficients in QCD Matter", Nucl. Phys. A **979**, edited by R. Rapp et al., 21–86 (2018).

[55] L. Betev, "Definition of the ALICE coordinate system and basic rules for sub-detector components numbering", tech. rep. ALICE-INT-2003-038 (ALICE Collaboration, 2003).

[56] C. Bierlich et al., "A comprehensive guide to the physics and usage of PYTHIA 8.3", HTML manual: `https://pythia.org/`, (Mar. 2022)

[57] J.-P. Blaizot and Y. Mehtar-Tani, "Jet Structure in Heavy Ion Collisions", Int. J. Mod. Phys. E **24**, 1530012 (2015).

[58] E. D. Bloom et al., "High-Energy Inelastic $e - p$ Scattering at 6° and 10°", Phys. Rev. Lett. **23**, 930–934 (1969).

[59] G. Bohm and G. Zech, "Introduction to Statistics and Data Analysis for Physicists; 3rd revised" (Verlag Deutsches Elektronen-Synchrotron, Hamburg, 2017), p. 488.

[60] P. Braun-Munzinger and B. Dönigus, "Loosely-bound objects produced in nuclear collisions at the LHC", Nucl. Phys. A **987**, 144–201 (2019).

[61] M. Breidenbach et al., "Observed Behavior of Highly Inelastic Electron-Proton Scattering", Phys. Rev. Lett. **23**, 935–939 (1969).

[62] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework", Nucl. Instrum. Meth. A **389**, 81–86 (1997).

[63] R Brun et al., "GEANT 3: user's guide Geant 3.10, Geant 3.11; rev. version" (CERN, Geneva, 1987).

[64] A. Buckley et al., "General-purpose event generators for LHC physics", Phys. Rept. **504**, 145–233 (2011).

[65] W. Busza, K. Rajagopal, and W. van der Schee, "Heavy Ion Collisions: The Big Picture, and the Big Questions", Ann. Rev. Nucl. Part. Sci. **68**, 339–376 (2018).

[66] A. Bzdak et al., "Mapping the Phases of Quantum Chromodynamics with Beam Energy Scan", Phys. Rept. **853**, 1–87 (2020).

[67] M. Cacciari and G. P. Salam, "Dispelling the N^3 myth for the k_t jet-finder", Phys. Lett. B **641**, 57–61 (2006).

[68] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", JHEP **04**, 063 (2008).

[69] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual", Eur. Phys. J. C **72**, 1896 (2012).

[70] S. Carrazza, S. Forte, and J. Rojo, "Parton Distributions and Event Generators", in 43rd International Symposium on Multiparticle Dynamics (2013), pp. 89–96.

[71] CBM Collaboration, "The Transition Radiation Detector of the CBM Experiment at FAIR: Technical Design Report for the CBM Transition Radiation Detector (TRD)", tech. rep. FAIR Technical Design Report (CBM Collaboration, Darmstadt, 2018), 165 p.

[72] S. Chatrchyan et al., "The CMS Experiment at the CERN LHC", JINST **3**, S08004 (2008).

[73] S. Chatrchyan et al., "Inclusive b-jet production in pp collisions at $\sqrt{s} = 7$ TeV", JHEP **04**, 084 (2012).

[74] S. Chatrchyan et al., "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", Phys. Lett. B **716**, 30–61 (2012).

[75] S. Chatrchyan et al., "Identification of b-Quark Jets with the CMS Experiment", JINST **8**, P04013 (2013).

[76] J. C. Collins, D. E. Soper, and G. F. Sterman, "Factorization for Short Distance Hadron - Hadron Scattering", Nucl. Phys. B **261**, 104–142 (1985).

[77] P. Connor, "Precision measurement of the inclusive b jet production in proton-proton collisions with the CMS experiment at the LHC at $\sqrt{s} = 13$ TeV", PhD thesis (Universität Hamburg, 2018).

[78] R. Corke and T. Sjöstrand, "Interleaved Parton Showers and Tuning Prospects", JHEP **03**, 032 (2011).

[79] P Cortese, "ALICE transition-radiation detector: Technical Design Report", Technical design report. ALICE (CERN, Geneva, 2001).

[80] P Cortese et al., "ALICE forward detectors: FMD, TO and VO: Technical Design Report", Technical design report. ALICE, Submitted on 10 Sep 2004 (CERN, Geneva, 2004).

[81] P. Cortese et al., "ALICE electromagnetic calorimeter technical design report", tech. rep. (ALICE Collaboration, Sept. 2008).

[82] L. Cunqueiro and M. Płoskoń, "Searching for the dead cone effects with iterative declustering of heavy-flavor jets", Phys. Rev. D **99**, 074027 (2019).

[83] G. D'Agostini, "A multidimensional unfolding method based on Bayes' Theorem", tech. rep. (DESY, Hamburg, 1994).

[84] H Damerau et al., "LHC Injectors Upgrade, Technical Design Report" (2014).

[85] G. Dellacasa et al., "ALICE technical design report of the inner tracking system (ITS)", tech. rep. (ALICE Collaboration, June 1999).

[86] K. Demmich, "Reconstruction of Beauty Jets in Proton-Proton Collisions at 13 TeV with ALICE", Analysis Note 1233 (ALICE Collaboration, 2022).

[87] Y. L. Dokshitzer, "Perturbative QCD for beginners", (1996) `http://cds.cern.ch/record/312932`.

[88] Y. Dokshitzer, V Khoze, and S. Troyan, "On specific QCD properties of heavy quark fragmentation ('dead cone')", Journal of Physics G: Nuclear and Particle Physics **17**, 1602 (1999).

[89] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.", Sov. Phys. JETP **46**, 641–653 (1977).

[90] F. A. Dreyer, G. P. Salam, and G. Soyez, "The Lund Jet Plane", JHEP **12**, 064 (2018).

[91] E. Epple and J. Mulligan, "Inclusive full jet measurements in pp collisions at $\sqrt{s}_{\mathrm{NN}} = 5.02\,\mathrm{TeV}$ with ALICE", Analysis Note 852 (ALICE Collaboration, 2019).

[92] L. R. Evans and P. Bryant, "LHC Machine", JINST **3**, This report is an abridged version of the LHC Design Report (CERN-2004-003), S08001. 164 p (2008).

[93] L. Feldkamp, "Reconstruction of Beauty Jets in Proton-Proton Collisions at $\sqrt{s} = 7\,\mathrm{TeV}$ with ALICE", PhD thesis (Westfälische Wilhelms-Universität Münster, Münster, 2018).

[94] N. Fischer et al., "Vincia for Hadron Colliders", Eur. Phys. J. C **76**, 589 (2016).

[95] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", Sept. 2007.

[96] S. Frixione, P. Nason, and B. R. Webber, "Matching NLO QCD and parton showers in heavy flavor production", JHEP **08**, 007 (2003).

[97] R. Fruhwirth, "Application of Kalman filtering to track and vertex fitting", Nucl. Instrum. Meth. A **262**, 444–450 (1987).

[98] "Future high-energy pp programme with ALICE", tech. rep. (ALICE Collaboration, 2020).

[99] R. Glauber, "High-Energy Collision Theory", Interscience Publishers **vol.1**, edited by W. Brittin and L. Dunham, 315 (1959).

[100] V. N. Gribov and L. N. Lipatov, "Deep inelastic e p scattering in perturbation theory", Sov. J. Nucl. Phys. **15**, 438–450 (1972).

[101] J. Haffner, "The CERN accelerator complex. Complexe des accélérateurs du CERN", General Photo (2013).

[102] H. Hassan, "Measuring the b-jet cross section in pp collisions at $\sqrt{s} = 5.02\,\mathrm{TeV}$", Analysis Note 982 (ALICE Collaboration, 2019).

[103] L. B. Havener, "Jet splitting measurements in Pb-Pb and pp collisions at 5.02 TeV with ALICE", 28th International Conference on Ultrarelativistic Nucleus-Nucleus Collisions, Wuhan, 2019.

[104] U. W. Heinz and M. Jacob, "Evidence for a new state of matter: An Assessment of the results from the CERN lead beam program", (Jan. 2000)

[105] S. Höche, "Introduction to parton-shower event generators", in Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (2015), pp. 235–295.

[106] A. Hocker and V. Kartvelishvili, "SVD approach to data unfolding", Nucl. Instrum. Meth. A **372**, 469–481 (1996).

[107] P. Ilten et al., "Disentangling Heavy Flavor at Colliders", Phys. Rev. D **96**, 054019 (2017).

[108] "Jet Quenching Tools", accessed on 21.09.2022, `https://jetquenchingtools.github.io`.

[109] F. Jiménez, "Charm and bottom production at particle colliders", MA thesis (Lund University, 2014).

[110] T. M. Karbach and M. Schlupp, "Constraints on Yield Parameters in Extended Maximum Likelihood Fits", (Oct. 2012)

[111] V. Kartvelishvili, "Unfolding with Singular Value Decomposition", accessed on 19.09.22, `https://indico.cern.ch/event/107747/contributions/32647/attachments/24321/35005/kartvelishvili.pdf`.

[112] V. Khachatryan et al., "Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV", JHEP **03**, 136 (2011).

[113] V. Khachatryan et al., "Event generator tunes obtained from underlying event and multiparton scattering measurements", Eur. Phys. J. C **76**, 155 (2016).

[114] J. Klein, "Near- and mit-term future of experiments at the LHC", 29th International Conference on Ultrarelativistic Nucleus-Nucleus Collisions, Krakow, Apr. 2022.

[115] G. Landsberg, "LHC: Past, Present, and Future", in 25th Rencontres de Blois on Particle Physics and Cosmology (Sept. 2013).

[116] D. Lohner, "Anisotropic flow of direct photons in Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 2.76\,\mathrm{TeV}$", PhD thesis (Ruperto-Carola-University, Heidelberg, 2013).

[117] A. Lopes and M. L. Perrey, "FAQ-LHC The guide", `https://cds.cern.ch/record/2809109`, accessed on 19.09.22, 2022.

[118] A. Maire, "Production des baryons multi-étranges au LHC dans les collisions proton-proton avec l'expérience ALICE", PhD thesis (Université de Strasbourg, Sept. 2011).

[119] A. Maire, "Topological selections for V0 (K0s, Lambda) and Cascade (Xi, Omega) reconstruction in ALICE", General Photo, 2011.

[120] A. Maire, "Track reconstruction principle in ALICE for LHC run I and run II. Principes de reconstruction de traces dans ALICE pour les runs I et II du LHC", General Photo, 2011.

[121] F. Maltoni, M. Selvaggi, and J. Thaler, "Exposing the dead cone effect with jet substructure techniques", Phys. Rev. D **94**, 054015 (2016).

[122] V. I. Man'ko et al., "ALICE Photon Spectrometer (PHOS): Technical Design Report", Technical design report. ALICE (CERN, Geneva, 1999).

[123] "Measurement of the nuclear modification factor of b-jets in 5.02 TeV Pb+Pb collisions with the ATLAS detector", ATLAS Collaboration, (Apr. 2022)

[124] "Measurements of b jet shapes in PbPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV", tech. rep. (CERN, Geneva, 2022).

[125] L. C. Mendez, "Jet production and suppression in heavy-ion collisions", 29th International Conference on Ultrarelativistic Nucleus-Nucleus Collisions, Krakow, 2022.

[126] E. Mobs, "The CERN accelerator complex in 2019. Complexe des accélérateurs du CERN en 2019", General Photo (2019).

[127] B. Müller, "Diagnosing the Quark-Gluon Plasma", (June 2021)

[128] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms", JHEP **11**, 040 (2004).

[129] "New State of Matter created at CERN", accessed on 19.09.22, CERN Press Release (2000).

[130] E. Norrbin and T. Sjostrand, "QCD radiation off heavy particles", Nucl. Phys. B **603**, 297–342 (2001).

[131] J. Park and V. Singh, "Measurement of electrons from beauty-hadron decays in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with ALICE", 29th International Conference on Ultrarelativistic Nucleus-Nucleus Collisions, Krakow, 2022.

[132] F. Piuz et al., "ALICE high-momentum particle identification: Technical Design Report", Technical design report. ALICE (CERN, Geneva, 1998).

[133] S. Plätzer, "Introduction to Event Generators", Mcnet Monte Carlo School, Prato, 2018.

[134]B. Povh et al., "Teilchen und Kerne – Eine Einführung in die physikalischen Konzepte" (Springer-Verlag Berlin Heidelberg, 2009).

[135]Prof. Dr. Christian Klein-Bösing, "Study of the Quark-Gluon Plasma with Hard and Electromagnetic Probes", Habilitation (Westfälische Wilhelms-Universität Münster, 2013).

[136]A. Purcell, "Go on a particle quest at the first CERN webfest. Le premier webfest du CERN se lance à la conquête des particules", (2012) `https://cds.cern.ch/record/1473657`.

[137]F. Reidt, "Upgrade of the ALICE ITS detector", Nucl. Instrum. Meth. A **1032**, 166632 (2022).

[138]M. Riordan, "The Discovery of Quarks", Science **256**, 1287–1293 (1992).

[139]C. Grojean and M. Spiropulu, eds., "Elements of QCD for hadron colliders", Comments: 7 lectures, 234 pages, published as CERN Yellow Report http://cdsweb.cern.ch/record/1119304, CERN (CERN, Geneva, 2010).

[140]G. Salam, "Jets and jet substructure 2: using jets", Center for Future High Energy Physics, Bejing, 2014.

[141]A. M. Sirunyan et al., "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", JINST **13**, P05011 (2018).

[142]A. M. Sirunyan et al., "Measurement of nuclear modification factors of Υ(1S), Υ(2S), and Υ(3S) mesons in PbPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV", Phys. Lett. B **790**, 270–293 (2019).

[143]T. Sjostrand and P. Z. Skands, "Transverse-momentum-ordered showers and interleaved multiple interactions", Eur. Phys. J. C **39**, 129–154 (2005).

[144]T. Sjöstrand, "On the $g \to Q\overline{Q}$ Rate", Dept. of Astronomy and Theoretical Physics, Lund University, 2014.

[145]T. Sjostrand, `https://pythia.org/latest-manual/examples/main08.html`, accessed on 02.09.22, 2022.

[146]T. Sjöstrand et al., "An introduction to PYTHIA 8.2", Comput. Phys. Commun. **191**, 159–177 (2015).

[147]P. Skands, "Introduction to QCD", in Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales (2013), pp. 341–420.

[148]P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 Tune", Eur. Phys. J. C **74**, 3024 (2014).

[149]L. A. Tarasovicova, "Di-hadron correlations of identified particles at high pT in pp collisions at the LHC", Presented 28 Apr 2022, PhD thesis (Universität Münster, 2022).

[150]A. Tauro, "ALICE Schematics", General Photo, 2017.

[151]"The ALICE definition of primary particles", ALICE Collaboration, (2017) `https://cds.cern.ch/record/2270008`.

[152]"The third run of the Large Hadron Collider has successfully started", accessed on 19.09.2022, CERN Press Release, 2022.

[153]M. Thomson, "Modern Particle Physics" (Cambridge University Press, United Kingdom, 2013).

[154]"Two circulating beams bring first collisions in the LHC", note = "accessed on 19.09.2022, CERN Press Release", 2009.

[155]W. Verkerke and D. Kirkby, "RooFit Users Manual v2.91" (2008).

[156]W. Verkerke, "Parameter estimation, $\mathcal{X}^2$ and likelihood", `https://www.physik.hu-berlin.de/de/gk1504/block-courses/autumn-2010/program_and_talks/Verkerke_part1/view`, accessed on 19.09.22, 2010.

Bibliography

[157] W. Verkerke and D. Kirkby, "The RooFit toolkit for data modeling", 2003.

[158] K. Yagi, T. Hatsuda, and Y. Miake, "Quark-Gluon Plasma", Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology (Cambridge University Press, 2008).

[159] N. Zardoshti, "First Direct Observation of the Dead-Cone Effect", Nucl. Phys. A **1005**, edited by F. Liu et al., 121905 (2021).

[160] P. A. Zyla et al., "Review of Particle Physics", PTEP **2020**, 083C01 (2020).

A. Additional information on the b-jet identification

A.1. Fit function for the description of the impact-parameter distributions

For several applications in this thesis, e.g. for the calculation of the jet probability as described in Sect. 5.4.1, smooth Sd_{xy} distributions are required as input. In these cases, each Sd_{xy} distribution is fit simultaneously with a combination of three functions: one for the central peak and two for the tails on both sides. The exponential of a Gaussian function

$$f_{\text{Gaus}}(x) = \exp\left(\frac{g_1}{2\pi g_3^2} \exp\left(-\frac{(x-g_2)^2}{g_3^2}\right)\right) \tag{A.1.1}$$

is applied for the peak around zero with $x_\alpha \le x \le x_\beta$. For the tails, a function of the kind

$$f_{\gamma,\text{Tail}} = t_{1,\gamma} \exp\left(t_{\gamma,2}(x-x_\gamma) + t_{\gamma,3}(x-x_\gamma)^2 + t_{\gamma,4}\exp\left(t_{\gamma,5}(x-x_\gamma)\right) - 1\right) \tag{A.1.2}$$

is utilised with $\gamma \in [\alpha, \beta]$. Thereby, the function $f_{\alpha,\text{Tail}}$ is defined for the left tail with $x \le x_\alpha$ and the function $f_{\beta,\text{Tail}}$ for the right tail with $x \ge x_\beta$. The functions $f_{\gamma,\text{Tail}}$ are forced to match f_{Gaus} at the transitions x_α and x_β from the peak to the tails, i.e.

$$f_{\text{Gaus}}(x_\alpha) = f_{\alpha,\text{Tail}}(x_\alpha) \quad \text{and} \quad f_{\text{Gaus}}(x_\beta) = f_{\beta,\text{Tail}}(x_\beta). \tag{A.1.3}$$

Thus, the combined function is continuous over the complete fit range.

A.2. Sd_{xy} **distributions**
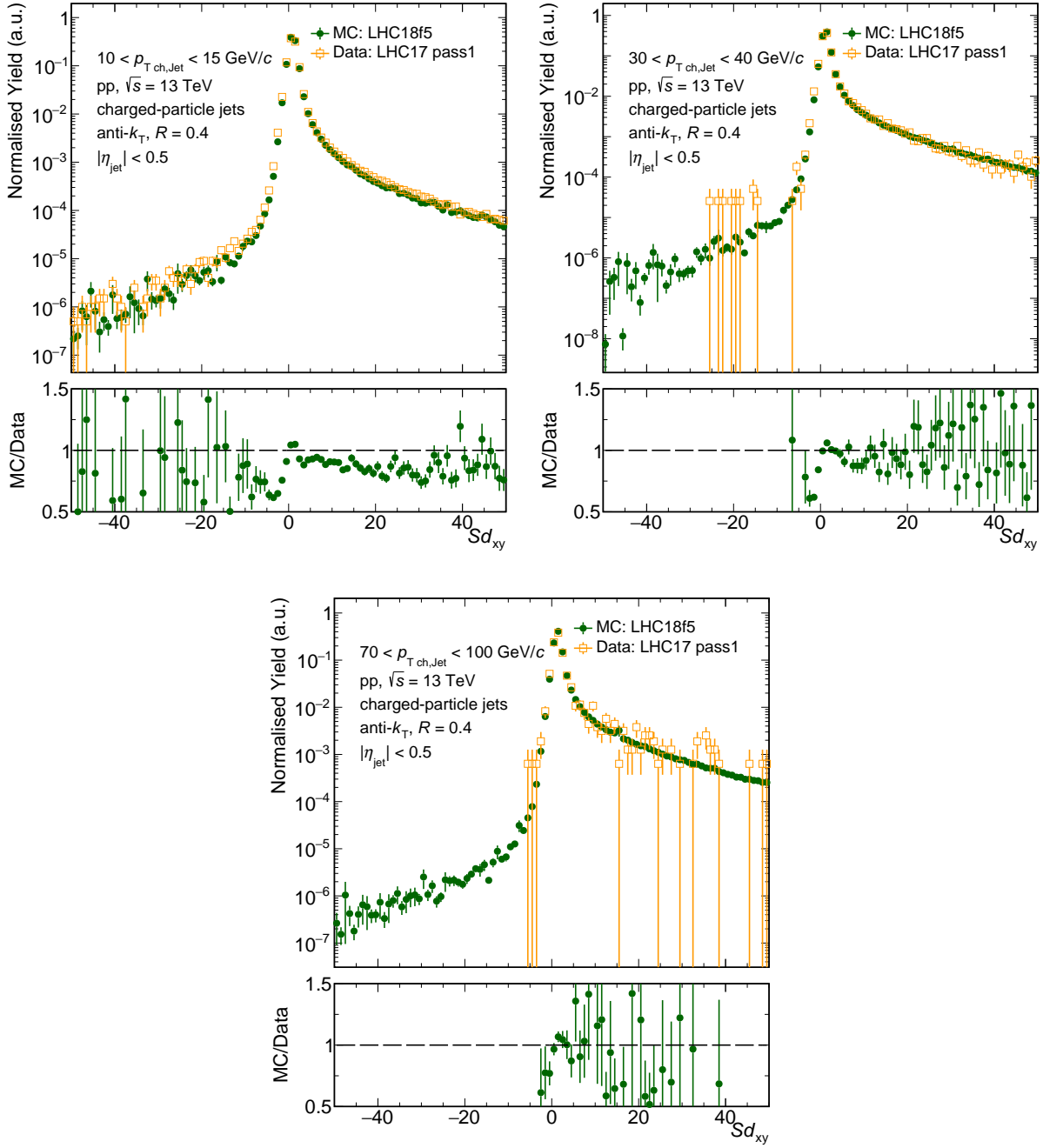
Figure A.1. – Comparison of the Sd_{xy} distributions for **N1-type tracks** within inclusive jets between data and MC at low, medium and high $p_{T,ch\ jet}$.
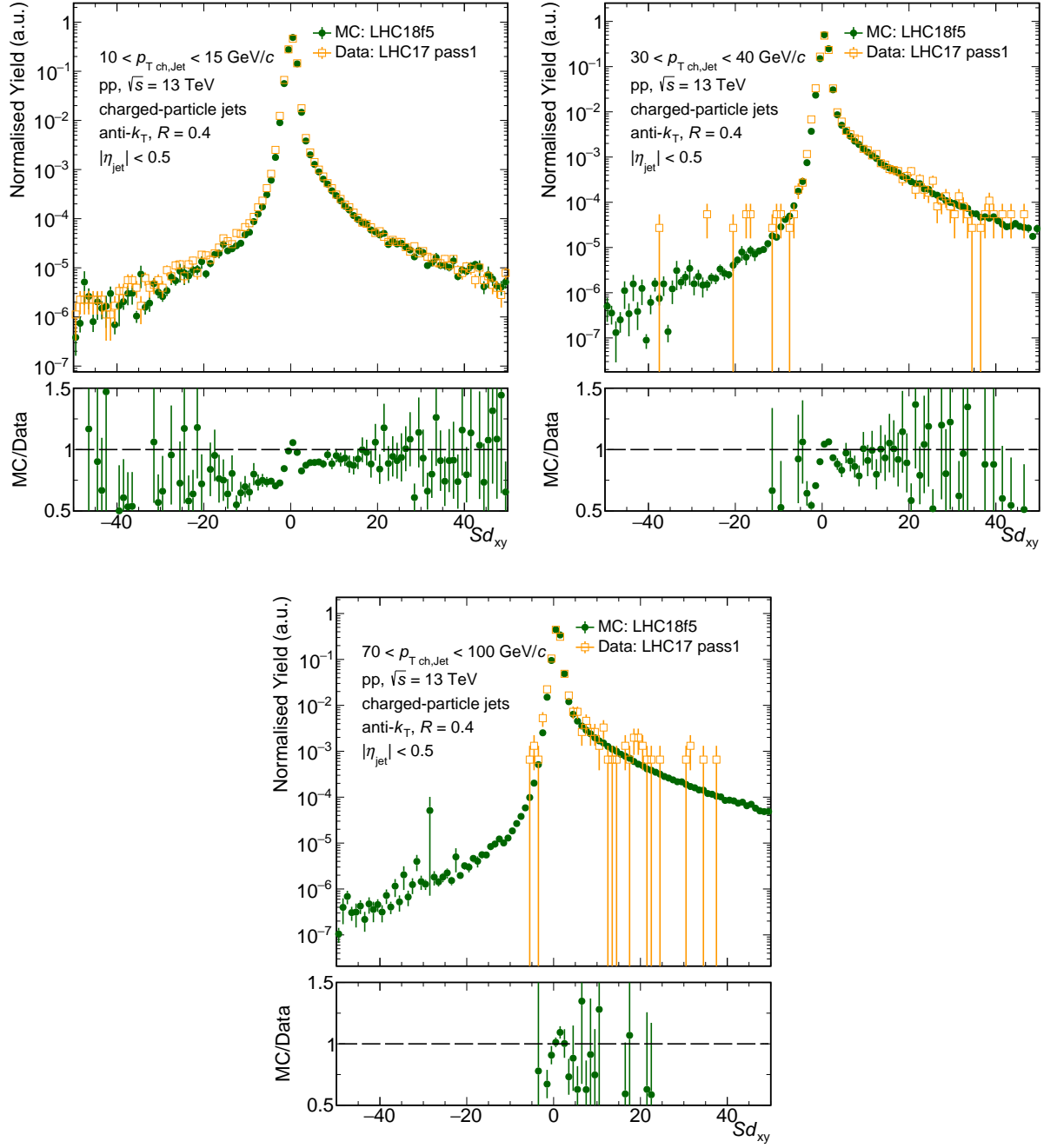
Figure A.2. – Comparison of the Sd_{xy} distributions for **N2-type tracks** within inclusive jets between data and MC at low, medium and high $p_{\mathrm{T,ch\ jet}}$.

B. Additional information on the b-jet tagging performance estimation

B.1. Parameter correlations assumed for the estimation of the statistical uncertainty

The number N^{Tag} of events tagged by the b-jet tagger per $p_{\mathrm{T,ch\ jet}}$ bin, the number $N^{\mathrm{Tag}}_{\mathrm{lnJP}}$ of tagged events with well-defined ln(JP) and the number N_{lnJP} of untagged events with well-defined ln(JP) are assumed to be fully correlated. For all other observable combinations, corresponding correlations are neglected. Following [59], the statistical uncertainty on the number of b-jets N_{b} is thus given by

$$\left(\frac{\Delta N^{\mathrm{b}}}{N^{\mathrm{b}}}\right)^2 = \text{Uncorrelated Uncertainty} - 2\frac{\Delta N_{\mathrm{lnJP}}}{N_{\mathrm{lnJP}}}\frac{\Delta N^{\mathrm{Tag}}_{\mathrm{lnJP}}}{N^{\mathrm{Tag}}_{\mathrm{lnJP}}} \tag{B.1.1}$$

$$- 2\frac{\Delta N^{\mathrm{Tag}}}{N^{\mathrm{Tag}}}\frac{\Delta N^{\mathrm{Tag}}_{\mathrm{lnJP}}}{N^{\mathrm{Tag}}_{\mathrm{lnJP}}} + 2\frac{\Delta N^{\mathrm{Tag}}}{N^{\mathrm{Tag}}}\frac{\Delta N_{\mathrm{lnJP}}}{N_{\mathrm{lnJP}}}. \tag{B.1.2}$$

The uncertainty on the b-jet tagging efficiency can be derived from Eqn. 5.4.5 and is given by

$$\left(\frac{\Delta \epsilon_{\mathrm{b}}}{\epsilon_{\mathrm{b}}}\right)^2 = \text{Uncorrelated Uncertainty} - 2\frac{\Delta N_{\mathrm{lnJP}}}{N_{\mathrm{lnJP}}}\frac{\Delta N^{\mathrm{Tag}}_{\mathrm{lnJP}}}{N^{\mathrm{Tag}}_{\mathrm{lnJP}}}. \tag{B.1.3}$$

Considering these parameter correlations has a negligible effect on the corresponding relative statistical uncertainties in comparison to considering only uncorrelated uncertainties. As it can be seen in Fig. B.1, the relative uncertainty at large $p_{\mathrm{T,ch\ jet}}$ is decreased by a few percent whereby no change is visible for low $p_{\mathrm{T,ch\ jet}}$.
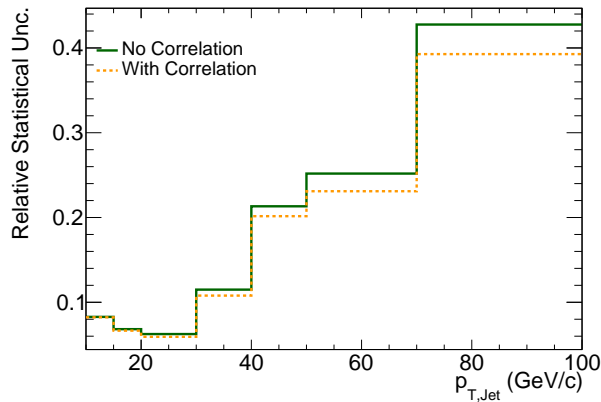

Figure B.1. – Comparison of the size of the statistical uncertainties on the performance-corrected b-jet spectrum for considering the observable correlations and for not doing so.

B.2. **Template fits for all** $p_{\text{T,ch jet}}$ **bins**
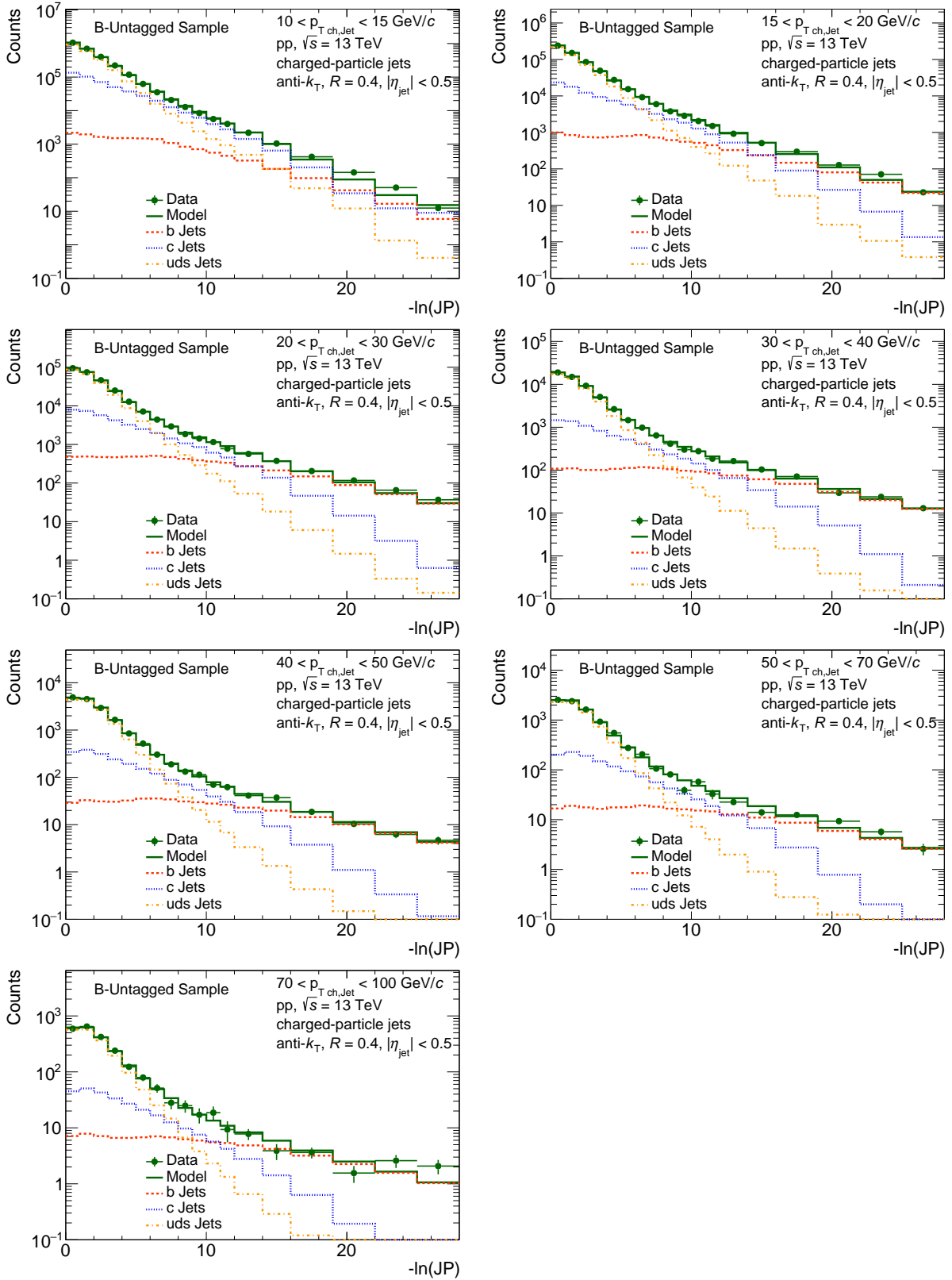

Figure B.2. – The fit templates, the fit model and the data for the performance estimation of the b-jet tagger. The results for fits to the default fit range are shown for all $p_{\text{T,ch jet}}$ bins of the **untagged data**. The fit templates are scaled with respect to their relative contribution to the fit model.
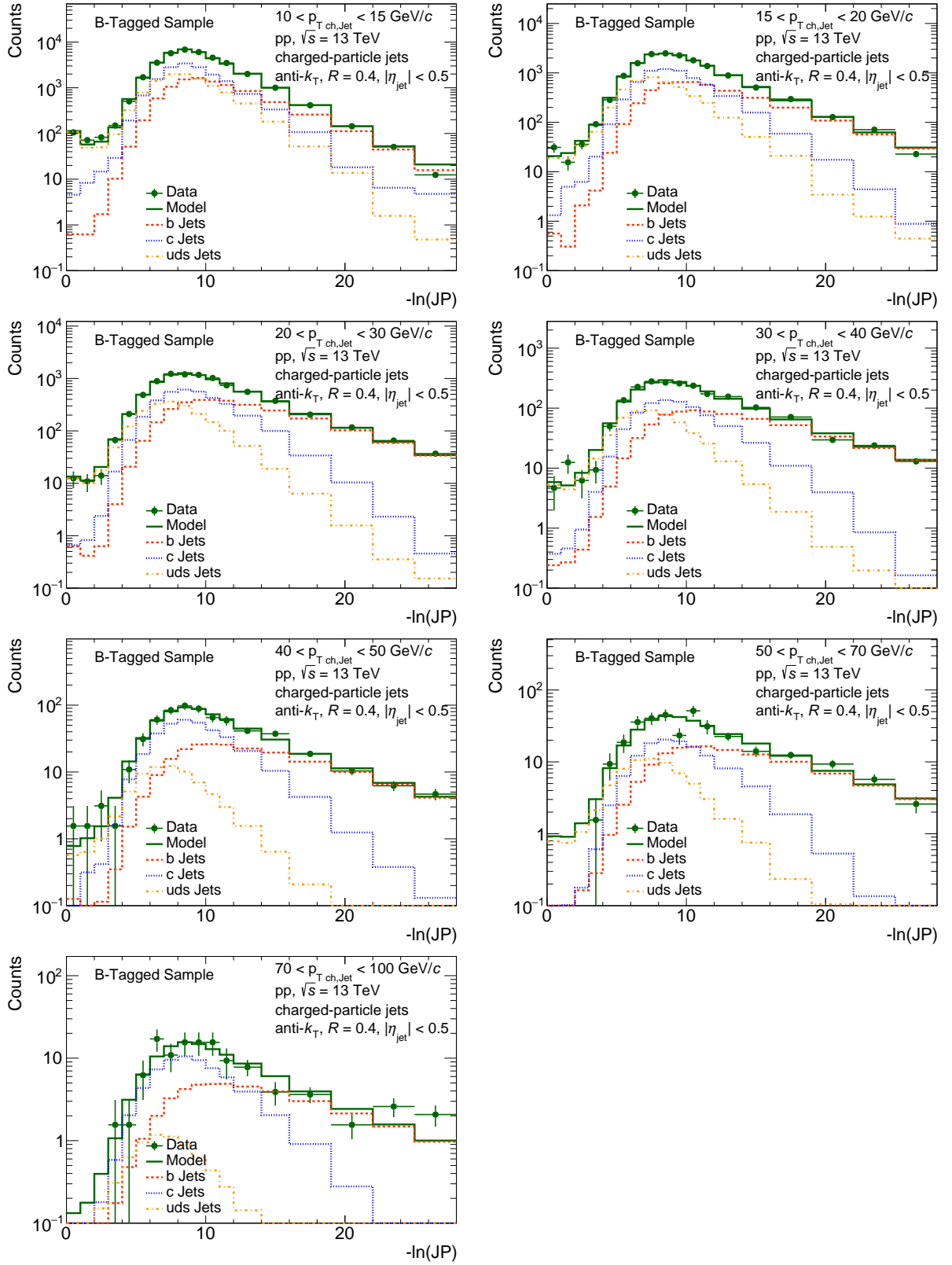
Figure B.3. – The fit templates, the fit model and the data for the performance estimation of the b-jet tagger. The results for fits to the default fit range are shown for all $p_{\mathrm{T,ch\ jet}}$ bins of the **tagged data**. The fit templates are scaled with respect to their relative contribution to the fit model.

C. Additional information on systematic uncertainties

C.1. Template-shape uncertainties

The $\ln(\mathrm{JP})$ Sampling Algorithm that is used for the estimation of the template-shape uncertainties in Sect. 5.6.1 utilises basic sampling concepts as described in [59, 133].

Hit-and-miss sampling   For this thesis, the algorithm that is used for the generation of a random number $x \in [a, b]$ according to a probability distribution $h(x)$ via hit-and-miss sampling and a maximum number of trials N_{trials} is:

> **for** N_{trials} **do**
> $\quad x \leftarrow \mathrm{rndm}()$ in $[a, b]$
> $\quad p \leftarrow \mathrm{rndm}()$ in $[0,1]$
> $\quad$ **if** $p < h(x)$ **then**
> $\quad\quad$ return x

Importance sampling   The algorithm that is used for the generation of a random number $y \in [a, b]$ according to a probability distribution $h(y)$ via importance sampling is:

> **for** N_{trials} **do**
> $\quad y \leftarrow \mathrm{rndm}()$ in $[a, b]$
> $\quad f \leftarrow g(y)$ with $g(y) > h(y)$
> $\quad p \leftarrow \mathrm{rndm}()$ in $[0,1]$
> $\quad$ **if** $p < f$ **then**
> $\quad\quad w \leftarrow \dfrac{h(y)}{g(y) \int g(y')\mathrm{d}y'}$
> $\quad\quad$ return w, y

Here, $g(y)$ is a histogram that encompasses the probability distribution $h(y)$ and is not necessarily normalised to unity. Based on the shape of $g(y)$, specific regions of the interval of interest can be oversampled with respect to others. This can help to improve the statistics of the final distribution of y in regions that otherwise suffer from low entry numbers. Every random number y contributes to corresponding histograms with a weight w that corrects for the oversampling.
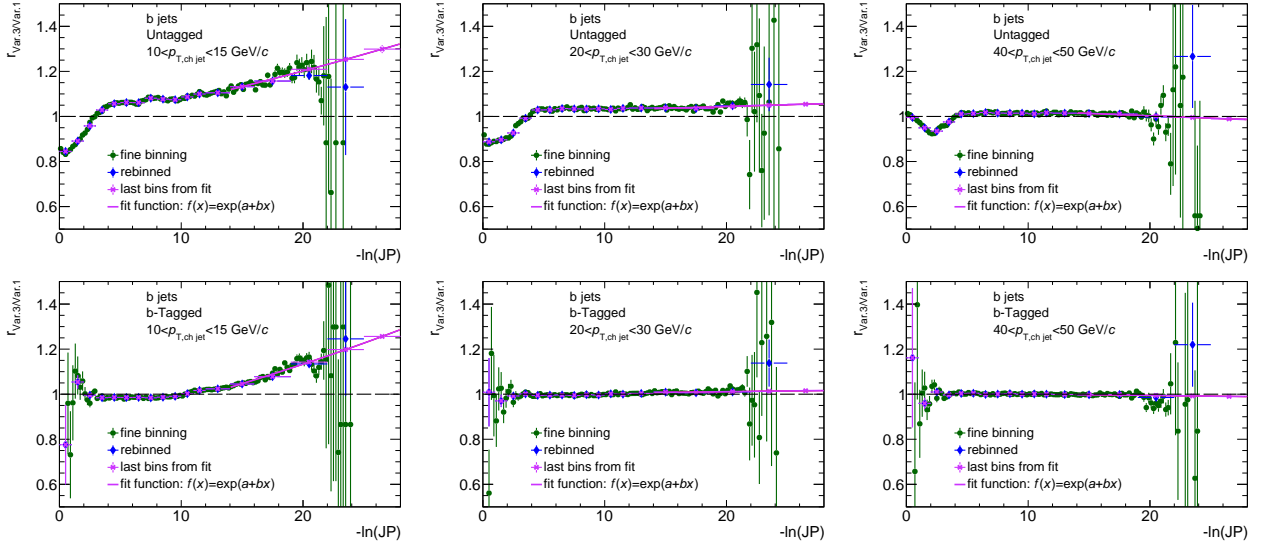
Figure C.1. – The ratios of the ln(JP) distributions for Var.3 and Var.1 as they are used for estimating the template shape uncertainties with respect to a change of the resolution function for **b jets** in the untagged (upper row) and tagged (lower row) data sample.
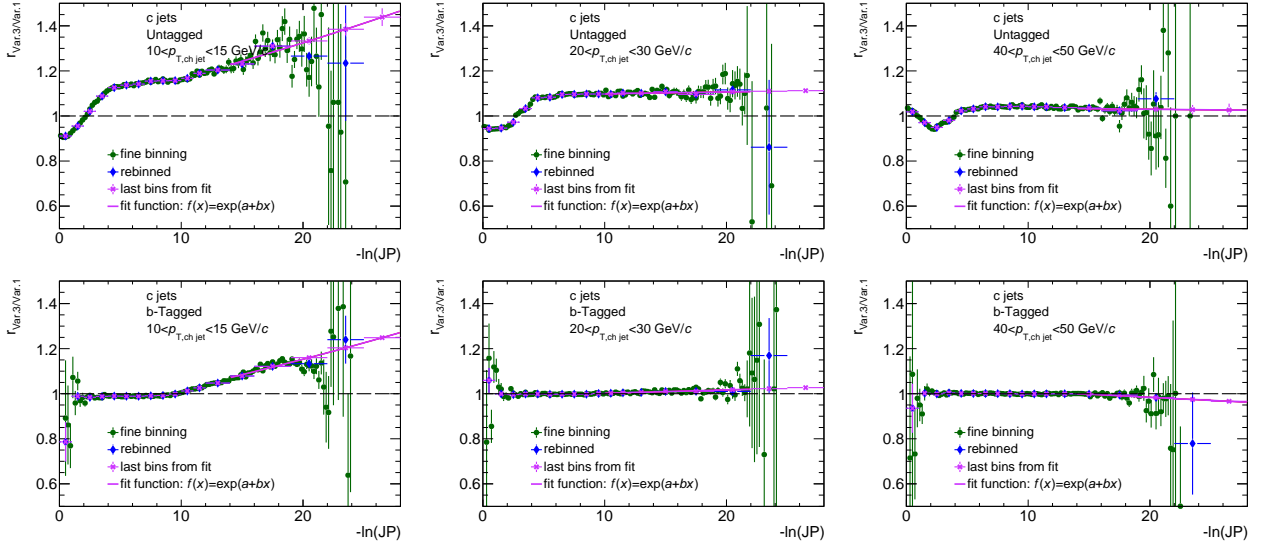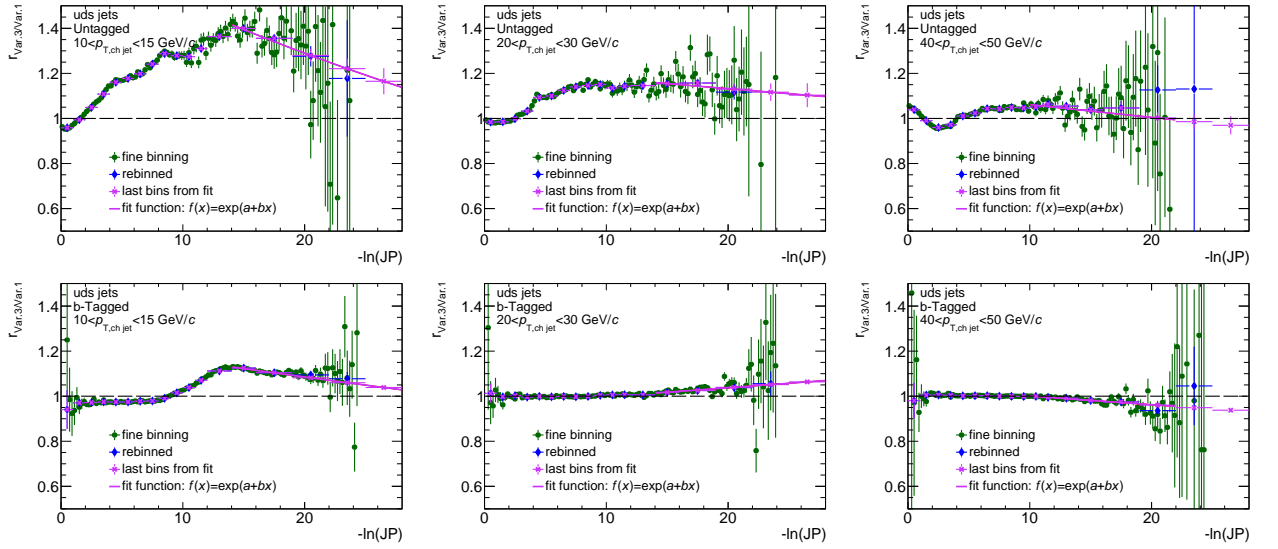

Figure C.2. – The ratios of the ln(JP) distributions for Var.3 and Var.1 as they are used for estimating the template shape uncertainties with respect to a change of the resolution function for **c jets** in the untagged (upper row) and tagged (lower row) data sample.

Figure C.3. – The ratios of the ln(JP) distributions for Var.3 and Var.1 as they are used for estimating the template shape uncertainties with respect to a change of the resolution function for **uds jets** in the untagged (upper row) and untagged (lower row) data sample.

C.2. V^0-jet uncertainty
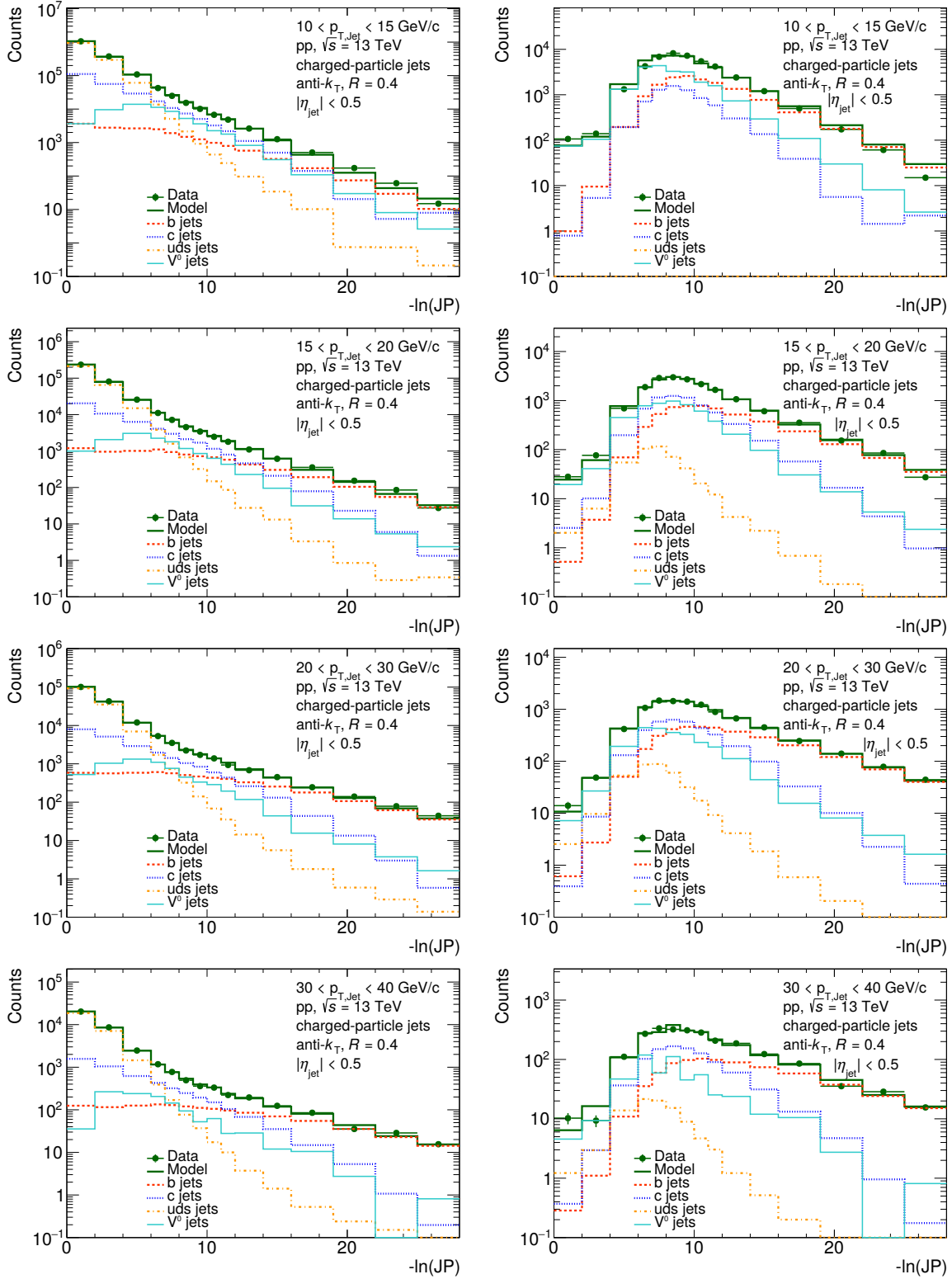
Figure C.4. – The fit templates, the fit model and the data for the performance estimation of the b-jet tagger with a dedicated V^0-jet template. The results for fits to the default fit range are shown for $p_{T,ch\ jet} < 40\,\mathrm{GeV}/c$ for the b-untagged data (left column) and the b-tagged data (right column). The fit templates are scaled with respect to their relative contribution to the fit model

C.3. Unfolding uncertainties – Choice of the algorithm

(a) Closure test I.

(b) Closure test II.

(c) Closure test III.

(d) Closure test IV.
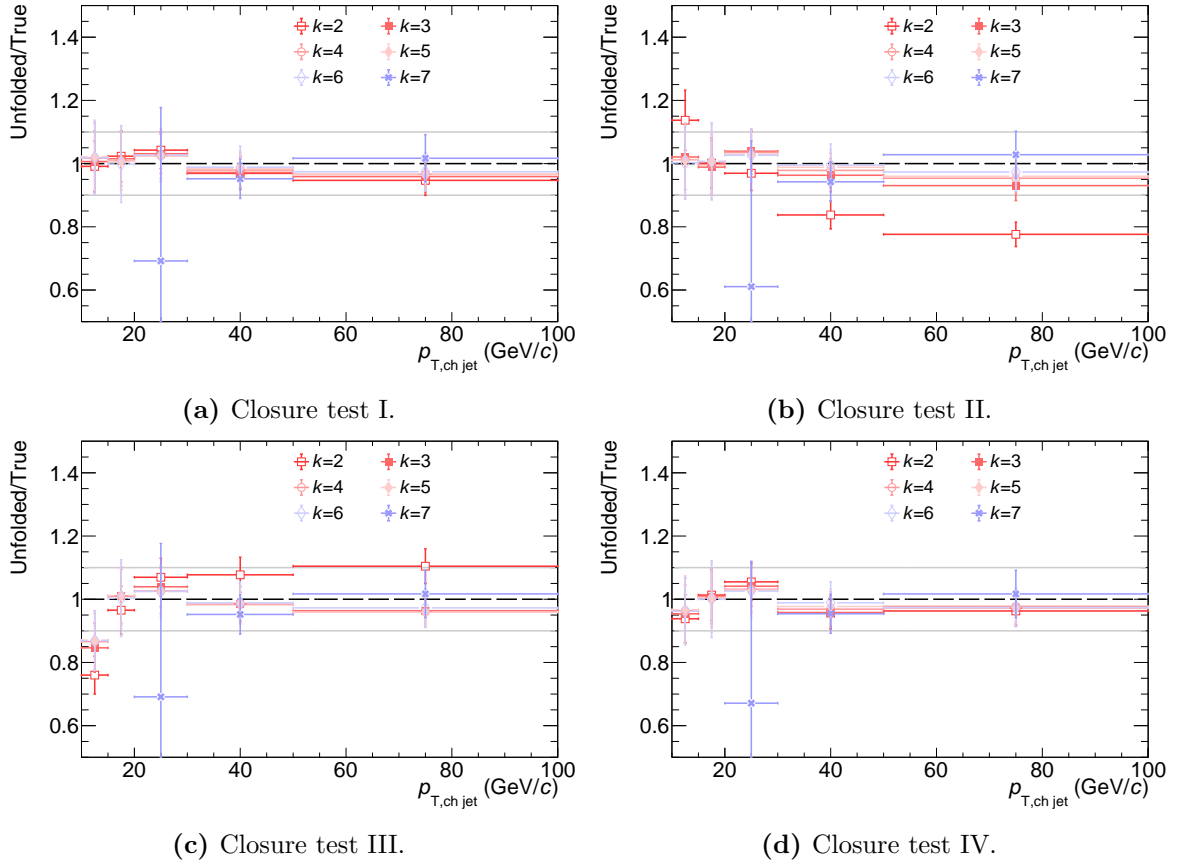
Figure C.5. – **a)-e)** The results of stability tests with respect to the SVD unfolding using MC simulations. **a)** The ratio of the unfolded pseudo data over the pseudo truth. **b)** As a) but with the pseudo data and pseudo truth scaled to match the original data. **c)** As a) but with the prior scaled to match the original data. **d)** As a) but with the prior scaled by the ratio of simulations for collisions with a centre-of-mass energy of $\sqrt{s} =5.02$ GeV for POWHEG+PYTHIA dijet over the same simulations with PYTHIA 8. See Sect. 5.5.3 for more details on the MC stability tests.
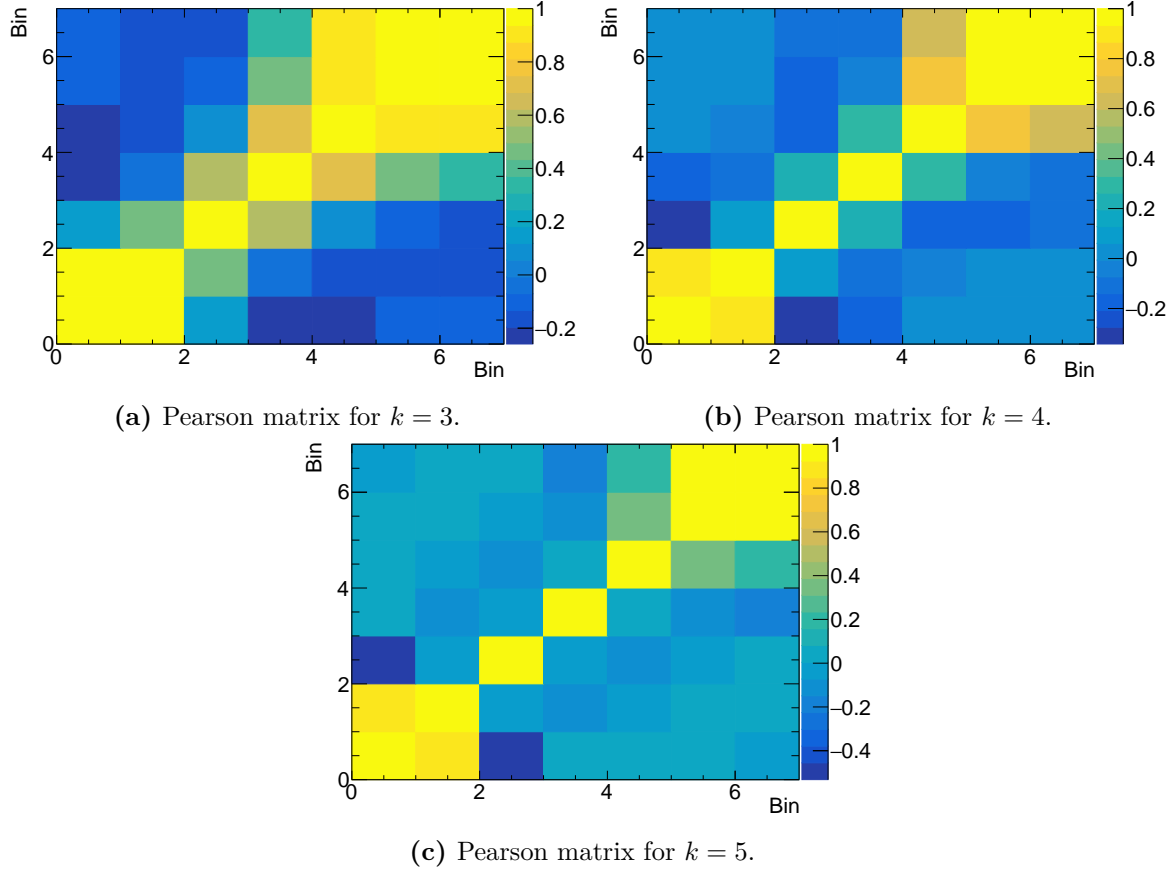
(a) Pearson matrix for $k = 3$.

(b) Pearson matrix for $k = 4$.

(c) Pearson matrix for $k = 5$.

Figure C.6. – The Pearson matrices for $k = 3, 4$ and 5 of the SVD unfolding procedure.

C.4. Tabular summary of uncertaintes

Table C.1. – The statistical and systematic uncertainties that are assumed for the final b-jet spectrum are provided for three $p_{\text{T,ch jet}}$ intervals. Thereby, the systematic uncertainties for the tracking resolution and the secondary vertex contamination have been adopted from [21].

$p_{\text{T,ch jet}}$	$10 - 15\,\text{GeV}/c$	$30 - 40\,\text{GeV}/c$	$70 - 100\,\text{GeV}/c$
Tracking efficiency	7.0	11.3	9.7
Tracking resolution [21]	0.4	0.1	3.7
Secondary vertex contamination [21]	1.6	2.3	4.0
Plateau region	6.2	0.6	2.3
Shape	8.0	3.9	4.1
Unfolding	14.2	4.4	3.5
pc cat.1: lower upper bound	11.8	4.0	6.4
pc cat.1: upper upper bound	10.5	4.6	4.6
Statistical uncertainty	7.9	11.7	37.8
Sys. lower bound	19.7	13.0	13.3
Sys. upper bound	19.0	13.1	12.5

C.5. Details on PYTHIA standalone simulations

For replicability of the PYTHIA 8 standalone simulation which is compared to the p_T-differential b-jet cross section in Sect. 5.7 as well as the simulation which is used for studying the Lund Plane in Sect. 6, the applied settings are listed in the following sections. For details on the mentioned PYTHIA settings it is referred to [56].

C.5.1. Reference for b-jet cross section

- version: PYTHIA 8.307 [56]

- tune: Monash 2013

- pdf: NNPDF2.3 LO, $\alpha_s(M_Z) = 0.130$ (default setting)

- fastjet version: 3.3.3

- The simulation has been done in 15 $\hat{p}_T$ bins with the following boundaries

$$0, 21, 28, 36, 45, 57, 70, 85, 99, 115, 132, 150, 169, 190, 212, 235.$$

  The first $\hat{p}_T$ bin $(0 < \hat{p}_T < 21\,\text{GeV}/c)$ has been simulated with the process flag Soft-QCD::nonDiffractive. For all other bins, the flag HardQCD::all has been applied. To avoid overlap between soft and hard events, those events simulated with SoftQCD::nonDiffractive are rejected if the simulated $\hat{p}_T$ is larger than the upper bin boundary as recommended in [145].

- In accordance with the ALICE jet finding framework [34], which clusters only physical primary particles (see Sect. 2.2.1), the daughter particles of the long-lived hadrons

$$K_s^0, \Lambda^0, \Sigma^\pm, \Xi^-, \Xi^0, \Omega^-$$

  that decay via the weak interaction have not been considered for the jet finding.

C.5.2. Lund-Plane simulation study

- PYTHIA version: PYTHIA 8.240 [146]

- tune: Monash 2013

- pdf: NNPDF2.3 LO, $\alpha_s(M_Z) = 0.130$ (default setting)

- fastjet version: 3.2.2

- The simulation has been done in 4 $\hat{p}_T$ bins with the following boundaries

$$20, 40, 60, 100, 200.$$

  All bins have been simulated with the process flag HardQCD::all.

Contributions to conferences

Talks

- 2022, International Conference on High-Energy Physics (06.07.-13.07., online/Bologna)
  "Beauty production in small systems with ALICE at the LHC "

- 2022, DPG-Frühjahrstagung (28.03.-01.04, online):
  "Reconstruction of beauty jets in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with ALICE"

- 2021, DPG-Herbsttagung (30.08.-03.09., online):
  "Reconstruction of Bottom Jets in Proton-Proton Collisions at $\sqrt{s} = 13\,\text{TeV}$ with ALICE"

- 2019, Graduiertenkolleg 2149 Annual Retreat (23.09.-26.09., Bramsche):
  "Investigating the Dead Cone Effect for Bottom Jets on the Lund Plane in pp Collisions at $\sqrt{s} = 13\,\text{TeV}$"

- 2019, ALICE Germany Meeting (09.09.-11.09., Seeon):
  "Further Results from HFE and Bottom-Jet Analyses"

- 2019, DPG-Frühjahrstagung (17.03.-22.03., München):
  "Reconstruction of Beauty Jets in Proton-Proton Collisions at $\sqrt{s} = 13\,\text{TeV}$ with ALICE"

- 2018, DPG-Frühjahrstagung (26.02.-02.03., Bochum):
  "Separation of Heavy-Flavour Production Mechanisms via Two-Particle Angular Correlations in Proton-Proton Collisions at $\sqrt{s} = 2.76\,\text{TeV}$"

Posters

- 2022, 29th International conference on ultrarelativistic nucleus-nucleus collisions (04.04.-10.04., online/Kraków)
  "Reconstruction of beauty jets in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with ALICE"

- 2020, LHCC meeting (19.02., CERN):
  "The Dead Cone Effect in Monte Carlo Simulations of the Lund Plane for Bottom Jets in Proton-Proton Collisions at $\sqrt{s} = 13\,\text{TeV}$"

- 2019, International Conference on Strangeness in Quark Matter (09.06.-15.06., Bari):
  "Reconstruction of Bottom Jets in Proton-Proton Collisions at $\sqrt{s} = 13\,\text{TeV}$ with ALICE"

Further contributions

- 2019, 28th International conference on ultrarelativistic nucleus-nucleus collisions (03.11.-09.11.,Wuhan):
  Contributed with simulations of projections of the Lund Plane for PbPb collisions at $\sqrt{s} = 5.02\,\text{TeV}$ in [35] to presentation in [103]. The differences between simulations with JEWEL and PYTHIA embedded in thermal background have been investigated with the Jet Quenching Tools framework [108].

Acknowledgements – Danksagungen

Eidesstattliche Erklärung

Ich versichere, dass ich nicht wegen eines Verbrechens verurteilt worden bin, zu dem ich meine wissenschaftliche Qualifikation missbraucht habe.

Münster, _____

Des weiteren versichere ich gemäß §7 Absatz 1(f) der Promotionsordnung des Fachbereiches Physik der Westfälischen Wilhelms-Universität vom 29.05.2019, dass ich die vorgelegte Dissertation selbständig und ohne unerlaubte Hilfe angefertigt habe, dass ich alle in Anspruch genommenen Quellen und Hilfsmittel in der Dissertation angegeben habe und die Dissertation nicht bereits anderweitig als Prüfungsleistung vorgelegen hat.

Münster, _____

Lebenslauf

Name	Katharina Demmich geb. Garner
Geburtsdatum	25.01.1992
Geburtsort	Bochum
Staatsangehörigkeit	Deutsch
Eltern	Marion Garner und Harald Garner

Studium

seit 2018	Promotionsstudium Physik
	Westfälische Wilhelms-Universität Münster
	Unter der Betreuung von Prof. Dr. Christian Klein-Bösing
2014 - 2017	Master of Science in Physik
	Westfälische Wilhelms-Universität Münster
März 2016 - Aug. 2017	Masterarbeit
	"Separation of Heavy-Flavour Production Mechanisms via Two-Particle Angular Correlations in pp Collisions at $\sqrt{s} = 2.76\,\text{TeV}$"
Okt. 2014 - Mar. 2015	Auslandssemester
	University of York (York, United Kingdom)
2011 - 2014	Bachelor of Science in Physik
	Westfälische Wilhelms-Universität Münster
Juli - Aug. 2014	Bachelorarbeit
	"Skalierungsverhalten identifizierter Teilchen in nuklearen Kollisionen vom RHIC zum LHC"

Schulausbildung

2002 - 2011	Allgemeine Hochschulreife
	Conrad-von-Soest-Gymnasium, Soest
1998 - 2002	Grundschule
	Bernhard-Honkamp-Schule, Welver

Berufstätigkeit

Jan. 2018 - Sep. 2022	Wissenschaftliche Mitarbeiterin am Institut für Kernphysik
	Westfälische Wilhelms-Universität Münster
Dez. 2015 - Dez. 2017	Studentische Hilfskraft am Institut für Kernphysik
	Westfälische Wilhelms-Universität Münster