Universiteit Antwerpen

Faculteit Wetenschappen

Departement Fysica

STUDY OF DOUBLE PARTON SCATTERING IN INCLUSIVE FOUR JET PRODUCTION IN PROTON-PROTON COLLISIONS AT LOW TRANSVERSE MOMENTUM AT $\sqrt{s} = 13$ TeV

Proefschrift voorgelegd tot het behalen van de graad van doctor in de wetenschappen:
fysica aan de Universiteit Antwerpen te verdedigen door

MAXIM PIETERS

Promotor:
Prof. dr Pierre Van Mechelen
Supervisor:
dr. Hans Van Haevermaet

Antwerpen, 2020

Abstract

A study of inclusive four jet production in proton-proton collisions at a centre-of-mass energy of 13 TeV is presented. The data sample was collected in 2016 with the CMS detector at the LHC during a low intensity run, with an integrated luminosity of 0.042 pb^{-1}. Differential cross sections are measured as a function of the jet transverse momentum, pseudorapidity, and several other observables that exploit angular correlations. The lowest jet transverse momentum cuts required in this paper are 35, 30, 25, and 20 GeV for the first, second, third and fourth leading jet respectively within $|\eta| < 4.7$, leading to a fiducial cross section of $\sigma = 2.77 \pm 0.02(\text{stat.})^{+0.68}_{-0.55}(\text{syst.})\mu\text{b}$. It is found that the data are very sensitive to different aspects of the underlying event, parton showers, and matrix element calculations. In particular the interplay between the de-correlations caused by parton showers and double parton scattering contributions is shown to be important. Models employing angular ordered parton showers, off-shell initial kinematics, as well as models with higher order matrix element calculations provide a better description of the data in certain observables, compared to standard leading-order models. The ΔS observable, which characterizes the azimuthal angular difference between the hard and soft jet pair, is used to extract a double parton scattering contribution by means of a template fit method. Model dependent values of sigma effective are calculated and compared to previous measurements. While all extracted values of the effective cross section show agreement with measurements performed at lower centre-of-mass energies, a strong model dependence of the double parton scattering contribution is found.

Acknowledgements

I would like to thank all the people who have supported me in any way or form during the past couple of years, people without whom the realization of this thesis would not have been possible or far harder.

First of all, I would like to thank Prof. Dr. Pierre Van Mechelen for granting me this opportunity to work in the field of high energy physics, for the many insights in the matter at hand, and for leading my research during these years. Second of all, thanks to Dr. Hans Van Haevermaet for having been my supervisor, for guiding me through the intricate web that is the CMS collaboration, for ever patiently answering any and all questions I had, and for the many fruitful discussions. Special thank go out to Prof. Dr. Sabine Van Doorslaer, Prof. Dr. Dirk Lamoen, Dr. Krzysztof Kutak, and Dr. Klaus Rabbertz for promptly accepting the roles of members of the jury of my dissertation.

Next, I would like to thank Sarah Van Mierlo for her kind assistance with any and all bureaucratic aspects of my Ph.D. Thank you to all the people in the CMS group at the University of Antwerp for creating a friendly, productive working environment and for the many pleasant, insightful, and necessary coffee breaks. Thanks to Jan Morez for the many moments of mental support and for allowing me to ventilate regarding unavoidable frustrations that have come with this thesis.

I would like to thank my father for stimulating my curiosity towards mathematics, physics and the understanding of the universe from early on in my life, and for being a person to whom I have looked up and will continue to look up to. Thanks to my mother for all the never ending love and support I have received throughout my life. I would also like to thank Charlotte, Maureen, Inge, and all the others in my ever supporting family.

Lastly, I would like to thank Giselle, Josse, Tim, Tom, Toni, and all friends who have been there for me over the years and who have helped relieve stress and who have helped me maintain my focus on the bigger picture.

Contents

List of Figures

List of Tables

Acronyms

ALICE A Large Ion Collider Experiment. 47

ATLAS A Toroidal LHC ApparatuS. 47

BBR Beam-Beam Remnants. 24

BFKL Balitsky-Fadin-Kuraev-Lipatov. 15

CCFM Ciafaloni-Catani-Fiorani-Marchesini. 16

CDM Color Dipole Model. 17

CERN Centre Européen pour la Recherche Nucléaire. 46

CFF Common FSQ Framework. 60

CHS Charged Hadron Subtraction. 63

CMS Compact Muon Solenoid. 46

CSC Cathode Strip Chambers. 55

DAQ Data Acquisition. 57

DGLAP Dokshitzer-Gribov-Lipatov-Altarelli-Parisi. 12

dPDF Double Parton Distribution Function. 29

DQM Data Quality Monitoring. 57

DT Drift Tube. 55

EB ECAL Barrel. 53

ECAL Electromagnetic CALorimeter. 53

EE ECAL Endcaps. 53

ES ECAL preShower dector. 53

FSR Final State Radiation. 23

HB HCAL Barrel. 53

HCAL Hadronic CALorimeter. 53

HE HCAL Endcaps. 53

HF HCAL Forward detector. 53

HLT High-Level Trigger. 57

HO HCAL Outer detector. 53

IP5 interaction point 5. 48

ISR Initial State Radiation. 23

JEC Jet Energy Correction. 62

JER Jet Energy Resolution. 150

JES Jet Energy Scale. 149

JSON JavaScript Object Notation. 57

LHC Large Hadron Collider. 46

LHCb Large Hadron Collider beauty. 47

LLA Leading Log Approximation. 12

LO Leading Order. 12

LV1 Level 1. 56

MB Minimum Bias. 85

MC Monte Carlo. 21

ME Matrix Element. 20

MET Missing Transverse Momentum. 63

miniAOD mini Analysis Object Data. 60

MPF Missing transverse momentum Projection Fraction. 70

MPI Multiple Partonic Interactions. 24

MRWTMD MRW-CT10nlo TMD. 38

NDF Number of Degrees of Freedom. 69

NLL Next-to-Leading Log. 35

NLO Next-to-Leading Order. 12

PBTMD PB-NLO-HERAI+II-2018-set2 TMD. 38

$PbWO_4$ lead tungstate. 53

PDF Parton Distribution Function. 10

PF Particle-Flow. 62

QCD Quantum Chromodynamics. 4

QED Quantum Electrodynamics. 9

QFT Quantum Field Theory. 4

QPM Quark Parton Model. 10

RPC Resistive Plate Chambers. 55

SM Standard Model. 4

SVD Singular Value Decomposition. 99

TEC Tracker EndCaps. 50

TIB Tracker Inner Barrel. 50

TID Tracker Inner Disks. 50

TMD Transverse Momentum Dependent Parton Distribution Function. 16

TOB Tracker Outer Barrel. 50

TOTEM ToTal Elastic and diffractive cross section Measurement. 47

TT Trigger Tower. 58

UE Underlying Event. 23

Chapter 1

Introduction

The standard model of physics (SM) [1,2] is a theory describing the elementary particles and three of the four fundamental interactions governing these particles in the universe. Although, the SM is incomplete, it is widely considered one of the most successful theories in physics due to its elegance and predictive power. One of the three interactions is the so-called "strong interaction". It is the interaction that holds the nuclei of atoms together. The interaction is, as the name implies, the strongest of the fundamental forces. Only elementary particles yielding a color charge or the quarks and gluons, participate in the strong interaction. The mathematical description is provided by quantum chromodynamics (QCD) and is studied through the measurement of high-energy proton-proton collisions by the Compact Muon Solenoid (CMS) detector at the Large Hadron Collider (LHC), operating at the Centre Européen pour la Recherche Nucléaire in Geneva. However, quarks and gluons can not be measured directly by any detector, due to the phenomenon that quarks and gluons cannot be isolated, also referred to as color confinement. As a direct consequence, clusters of colored particles or jets are observed in detectors rather than the individual quarks or gluons.

QCD provides a good description of jets with large transverse momentum (p_{T}) produced in high-energy proton-proton collisions. This is achieved by factorizing the cross section in a perturbatively calculable matrix element (ME) for the scattering between partons, on the one hand, and parton distribution functions (PDF) that describe the probability to find a parton within the proton, on the other hand. The latter contain information on the proton, which can not be perturbatively calculated, and are obtained by fitting available data. Measurements of the cross section for the production of inclusive high-p_{T} jets have been performed by the CMS collaboration at various centre-of-mass energies [3–5] and show good agreement with perturbative QCD predictions at next-to-leading order (NLO) accuracy. However, final states with multiple jets, and their correlations, call for further studies of the strong interaction.

Multijet final states can be produced in a single parton scattering (SPS). Depending on the order of the ME in the strong coupling factor, two or more jets can be produced in SPS. Radiation before and/or after the interaction between the partons, as described by parton shower models, can contribute additional jets to the final state. As such, predictions for multijet SPS processes provide an important test of the matching between fixed-order ME calculations and the parton shower formalism. An alternative approach, however, introduces a second hard scattering in the proton-proton collision, thus also contributing a number of jets to the final state. Such processes are in general referred to as double parton scattering (DPS), and represent the simplest case of multiple partonic interactions (MPI). The cross section of a DPS process, $\sigma^{\mathrm{DPS}}_{\mathrm{A,B}}$, where A and B denote two processes with their own respective cross sections σ_{A} and σ_{B}, can be expressed as:

$$\sigma^{\mathrm{DPS}}_{\mathrm{A,B}} = \frac{m}{2}\frac{\sigma_{\mathrm{A}} \cdot \sigma_{\mathrm{B}}}{\sigma_{\mathrm{eff}}}. \tag{1.1}$$

The factor m is a combinatorial factor, which is equal to one for identical processes and equal to two for non-identical processes. The effective cross section (σ_{eff}) reflects how strongly the occurrence of A and B is correlated. For fully uncorrelated production of A and B, σ_{eff} tends to the total inelastic cross section, while a small σ_{eff} indicates an enhanced simultaneous occurrence of A and B. For multijet production, SPS processes often exhibit strong kinematic correlations between all jets, while DPS processes will manifest a distinctly different behaviour. Indeed, the jets resulting from DPS are more often than not produced in two independent pairs in a back-to-back configuration. The DPS cross section rises with increasing centre-of-mass energy: at higher energy, smaller values of the proton's momentum fraction carried by the partons can be probed, resulting in a strong increase of the gluon density and accordingly a larger probability for DPS. A proper knowledge of DPS processes is therefore beneficial for a complete understanding of hadronic interactions.

Several DPS measurements at different centre-of-mass energies and for various final states have been performed. Studies including one or two photons in the final state have been published in [6–11]. Final states involving one or two vector bosons have been measured by the ATLAS and CMS collaborations [12–15]. Other studies have opted to involve the production of heavy flavours [16–20]. Older measurements in the four jet final state have been performed by the UA2 and CDF experiments [21, 22], while the ATLAS and CMS collaborations have more recently also performed DPS measurements with four jets [23–26]. The CMS collaboration additionally obtained a measurement for the final state with two b-tagged jets in combination with two light-flavoured jets [27].

This thesis presents an analysis of the inclusive production of four jets in proton-proton collisions at a centre-of-mass energy of 13 TeV. The data were collected with the CMS detector at the CERN LHC in 2016, during a run with a low probability for several proton-proton interactions occurring within the same or nearby bunch crossings (henceforward referred to as "pileup"). This allows not only to mitigate effects of such pileup, but also to reach down to low jet p_{T}, albeit necessitating a custom calibration of the jet energy scale. The collected data correspond to an integrated luminosity of 0.042 pb^{-1}. Data are corrected for detector effects by means of an unfolding procedure applied to the measured distributions.

Several aspects of multijet production are studied through the comparison of the distributions of "DPS-sensitive" observables predicted by various Monte Carlo (MC) models and obtained from data. These observables all exploit the differences in the kinematic correlations between the jets expected for SPS and DPS.

Two methods are used to extract the DPS cross section. A first estimate is obtained with a template method: a pure DPS signal sample is reconstructed from data by randomly mixing inclusive single jet events and is fitted along with several SPS-only background MC models to the distributions obtained from data. In an alternative method, SPS-only MC models are normalized to data in a DPS-free region and the DPS cross section is estimated from the differences between data and SPS-only MC models in regions where a DPS contribution is expected. Finally, the effective cross section is computed using eq. (1.1), with σ_A and σ_B measured from inclusive single jet production in data.

The thesis is organized as follows: a more extensive overview of the mathematical structure of the SM together with a more in-depth formulation of QCD is given in Chap. 2. In Chap. 3, the hadron collisions and all its components are described. After discussing the DPS pocket formula in detail, the chosen DPS-sensitive observables are defined and all models used to study the observables are detailed next. Finishing with a first simulation of all observables in order to justify the choices made. In Chap. 4, the whole infrastructure of the LHC and the CMS detector are discussed in detail. The data and simulated samples are specified in Chap. 5. The latter are used to correct the detector level data to generator level, on which the predictions are independent of any detector effects. Chap. 6 illustrates how different objects in proton-proton collisions are reconstructed. Of the necessary objects, jets still need to be calibrated through the sequential application of the jet energy corrections (JEC) factors, which had to be derived for this specific analysis. The event selection, defined in Chap. 7, is applied to the fully reconstructed and calibrated events. Next, the construction of a pure DPS sample from data, along with the methods of extraction of the σ_eff parameter is presented in Chap. 8. In Chap. 9, a full treatment of the detector effects is given. While the unfolding procedure, meant to correct for the detector effects, is outlined in Chap. 10, after which the detector level distributions are unfolded to obtain the generator level curves. However, a measurement is not free of observational errors. Certain steps in the analysis give rise to systematical uncertainties. All of them are treated in Chap. 11. Once the error treatment has been performed a comparative study between the data and different models is presented in Chap. 12, while the final conclusions are written down in Chap. 13.

Chapter 2

The Standard Model

In what follows the theoretical framework, together with core concepts and models needed to describe high-energy collisions, are introduced. We will start with a broad description of the Standard Model (SM) of physics. Afterward, the mathematical formulation of the strong nuclear force or Quantum Chromodynamics (QCD) will be discussed in more detail.

2.1 Introduction to the Standard Model

Throughout history continuous efforts have been made towards understanding the universe around us and towards formulating a fundamental and descriptive theory of all matter. Historically famous examples of such ideas and theories include the concept of atoms as indivisible building blocks of matter and the table of Mendeleev. The former was first proposed by Democritus in the fifth century BCE, while the latter was published in 1869. The strong ordering of the elements in the table of Mendeleev suggested the presence of an underlying and more fundamental substructure. During the 19th and the beginning of the 20th century, the substructure was probed and it was discovered that atoms are composed of a vacuum in which negatively charged electrons encompass a positive nucleus. The electrons, discovered by Thompson in 1898 [28], and nuclei are bound together by the electromagnetic force. Later on during the 1920s, concepts such as position and velocity had to be revisited due to the prominent role of the Heisenberg uncertainty principle [29] and the wave-particle duality of matter in the formulation of quantum mechanics. In 1932 [30], Chadwick discovered that the nuclei were not only composed of positively charged protons, but neutrally charged neutrons as well. They are "glued" together by the strong nuclear force, which effectively overcomes the electromagnetic repulsion between the positively charged protons. In turn the protons and neutrons were found to be consisting of quarks. A third force, the so-called weak nuclear force, allows for the decay of neutrons into protons accompanied by an electron through a process called β-decay. The framework of quantum mechanics, based on the Schrödinger equation [31], could not successfully describe the annihilation and creation of particles, which was necessary for processes like β-decay. A specific type of relativistically invariant gauge theories based on the concept of Quantum Field Theory (QFT), i.e. renormalizable Yang-Mills theories [32], resolved the problems existing in quantum mechanics. Further exploration of these theories along with different experiments led to the formulation of the SM during 1970-1973 [1,2].

The SM succeeds in describing three of the four fundamental forces, all except gravity, as well as classifying all elementary or fundamental particles, which are defined as particles with no further substructure. An overview of all known elementary particles is shown in Fig. 2.1. A first distinction between two main classes of elementary particles is made within the SM. On the one hand, fermions are particles with a half-integer spin. All leptons and quarks are fermions. Both groups of particles consist of three generations. The particles in each generation carry similar traits and properties, while their mass increases with each generation. On the other hand, bosons are elementary particles with an integer spin. The gauge bosons serve as the carriers of the three aforementioned fundamental forces. The Higgs boson, discovered in 2012 by the ATLAS and CMS collaborations [33,34], completes the SM. The Higgs mechanism was theorized as early as 1964 independently by François Englert and Robert Brout [35], and by Peter Higgs [36]. Its existence is essential to the SM, as without the Higgs mechanism massive particles can not exist within the framework of the SM in a consistent manner.

The framework of QFT [1,2] has provided the basis on which the SM is built. The Lagrangian approach has been adopted throughout the SM. All particle dynamics of the SM are contained in and describe by the Lagrangian function specific to the SM. The theoretical framework of QFT is a quantum mechanical and relativistically invariant formulation of classical field theory. In other words: each particle is treated as an excitation of its underlying quantum field, which permeates throughout all of space-time, and is described by a quantized wave function, where the whole framework is invariant under Lorentz transformations. QFT effectively combines quantum mechanics, classical field theory and special relativity [38]. Three types of quantum fields enter in the SM Lagrangian function. The fermions are described by spinor fields, while the force carrying bosons and the Higgs boson are represented by vector fields and a scalar field, respectively. A scalar field remains invariant under Lorentz transformations, while the vector fields transform in the same manner as a four-vector. Both obey the Klein-Gordon equation of motion [39,40]. The fermionic fields transform quite differently under Lorentz boosts and are represented by spinor fields. In contrast to the bosons, the fermions follow the Dirac equation of motion [41], opposed to the Klein-Gordon equation. A more detailed discussion, however, would take us too far from the topic at hand but can be found in [1,2].

The interactions governing the quarks and leptons, are the direct result of the symmetries imposed on the SM. The particles in the SM obey three local symmetries, i.e. the wave functions of particles remain invariant under local transformations of the Lie groups $U(1)_{\mathrm{EM}}$, $SU(2)$ and $SU(3)$ [1]. The subscript "EM" stands for the electromagnetic force. In order to preserve the invariance of the quark and lepton fields under the local transformations, additional gauge fields need to be included in the Lagrangian of the SM. The gauge fields correspond to the gauge bosons that mediate the three aforementioned fundamental forces. The photon γ is the force carrying boson of the electromagnetic interaction, which corresponds to the invariance of the SM under transformations of the $U(1)_{\mathrm{EM}}$ group. In the same manner, the gluon is connected to the strong nuclear force and the $SU(3)$ group. While the weak nuclear interaction is carried by the W^+, W^- and Z^0 bosons and exhibits invariance under transformations of the $SU(2)$ group. The elegance of the SM is highlighted here: the three fundamental forces and the force carrying

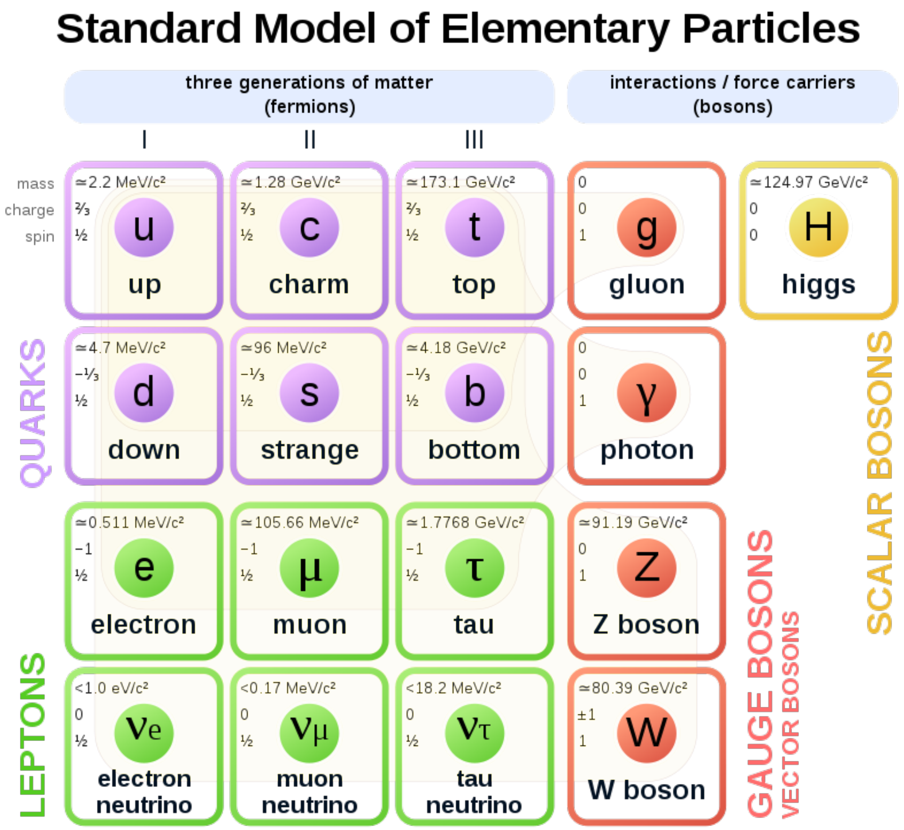


Figure 2.1: An overview of all the elementary particles of the SM, taken from [37]. Quarks, leptons and gauge bosons are shown in purple, green and red, respectively, while the Higgs boson is depicted in yellow. The mass, charge and spin of each particle are specified on the table.

gauge bosons described by the SM arise from the imposed local symmetries.

During the 1960s Sheldon Glashow observed similarities between the electromagnetic force and the weak nuclear force and succeeded in laying the foundation for the electroweak interaction [42]. Mohammed A. Salam and Steven Weinberg had the similar ideas [43, 44] and produced the same framework independently of Sheldon Glashow. The electroweak interaction is the unified description of the electromagnetic and the weak nuclear interaction, and is symmetric under transformations of the combined group $SU(2) \times U(1)_{\mathrm{Y}}$, where Y represents a newly conserved quantity, i.e. the weak hypercharge. A W bosons triplet (W^0, W^1 and W^2) and the boson singlet B^0 carry the electroweak interaction. However, the newly defined gauge fields do not correspond to the physical γ and Z^0 boson fields, they are a rotation of the non-physical W^0 and B^0 fields. The relation between the physical and non-physical fields is written as

$$\begin{pmatrix} |\gamma\rangle \\ |\mathrm{Z}^0\rangle \end{pmatrix} = \begin{pmatrix} \cos\theta_\mathrm{W} & \sin\theta_\mathrm{W} \\ -\sin\theta_\mathrm{W} & \cos\theta_\mathrm{W} \end{pmatrix} \begin{pmatrix} |\mathrm{B}^0\rangle \\ |\mathrm{W}^0\rangle \end{pmatrix} \tag{2.1}$$

Where θ_W is the weak mixing angle, also called the Weinberg angle, which has been determined through various experiments, e.g. the CMS collaboration found a value of $\sin^2\theta_\mathrm{W} = 0.23101 \pm 0.00053$ [45]. The remaining W^1 and W^2 bosons correspond with the earlier defined W^+ and W^- bosons. That being said, two issues in the description of the electroweak interaction still remained. On the one hand, the theory is not renormalizable, which is a requirement of the SM. Renormalization aims to treat infinities arising in perturbative calculations in a consistent manner. On the other hand, the four bosons of the electroweak interaction are massless since mass terms can not naively be added to the Lagrangian in a gauge invariant way. Both problems were resolved by spontaneous symmetry breaking and the Brout-Englert-Higgs mechanism.

The Brout-Englert-Higgs mechanism introduces the scalar Higgs field ϕ and an associated potential $V(\phi)$ to the SM. The shape of the potential $V(\phi)$ is quite peculiar, in the sense that it is shaped like a Mexican hat or sombrero, as depicted in Fig. 2.2. The form of the potential allows for the spontaneous breaking of its symmetry. The minimum of the potential is not centred at $V(0)$, the local maximum of the Mexican hat potential. Choosing a vacuum state, a state where the energy is minimal, results in the spontaneous breaking of the electroweak symmetry group $SU(2) \times U(1)_\mathrm{Y}$. The direction of the symmetry breaking can be chosen in such a way that the photon remains massless, while new gauge invariant mass terms for the Z^0, W^+ and W^- bosons appear in the Lagrangian. The remaining symmetry is described by the $U(1)_\mathrm{EM}$ group of the electromagnetic force.

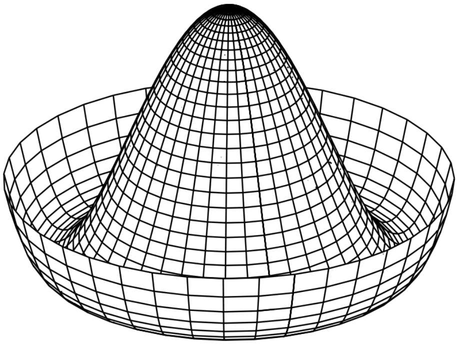


Figure 2.2: A schematic depiction of the potential $V(\phi)$, used in the spontaneous symmetry breaking framework and the Brout-Englert-Higgs mechanism, taken from [46]. By going from the point $V(0)$, which is the local maximum of the potential, to a vacuum state or minimum of the potential, the symmetry is spontaneously broken. The vacuum state can be freely chosen in such a way that the remaining symmetry is the the one described by the $U(1)_\mathrm{EM}$.

However, the story is not complete yet, as fermions are known to have masses as well, which are introduced in the SM Lagrangian through the inclusion of the Yukawa coupling

[47], named after Hideki Yukawa. As a new scalar field has been introduced to the Lagrangian of the SM, additional interactions between the Higgs field and all other fields should be added as long as they are not prohibited by the symmetries of the SM. The Yukawa coupling defines a gauge invariant interaction between a spinor field ψ and a scalar field ϕ. The interaction term brings forth a gauge invariant mass term for the fermions once the symmetry is spontaneously broken. The existence of the Higgs field saves the SM as a renormalizable Yang-Mills theory. The discovery of the Higgs boson in 2012 by the ATLAS and CMS collaborations, almost 50 years afters its theoretical conception, is therefore seen as a major triumph of the SM and will remain a landmark in the history of modern physics.

Apart from all the success the SM has had, it is not seen as the final theory. Within its framework the SM is consistent, but it still leaves major questions unanswered. One of the largest hurdling blocks of the SM is the exclusion of the fourth fundamental force. The SM model is incompatible with general relativity [48], the mathematical formulation of gravity. Candidate unifying theories have been proposed, e.g. supergravity, which combines principles from supersymmetry and general relativity [49]. However, effects of such theories have yet to be observed. Other unexplained phenomena include the matter-antimatter or so-called baryon imbalance [50], neutrino oscillations [51], the presence of dark matter [52] and dark energy [53] in the universe. As a last unexplained phenomenon we mention the hierarchy problem [54]. All aforementioned problems with the SM have been encountered by various experiments, however, anomalous behavior within the SM is bothersome as well. An example of the latter is the hierarchy problem. Earlier in the text, the renormalizability of the SM was mentioned as a means to deal with infinities in calculations. These infinities occur due to the fact that fundamental quantities, such as the mass of the Higgs boson, can have a different effective and fundamental value. The former is measured through experiments, while the latter is the quantity in the Lagrangian. The renormalization procedure connects both through the application of corrections to the fundamental values. Whenever a large discrepancy between the effective and fundamental values exists, an incredibly precise fine-tuning occurs between the fundamental value and the corrections in order to obtain the effective value. Such a precise fine-tuning does not seem natural. Arguments to justify the fine-tuning have been made in the sense of the anthropic principle, but border on philosophy rather than on physics. While theories beyond the SM aim to resolve the hierarchy problem through the formulation of underlying physical principle of which we have no knowledge as of yet.

The general framework of high-energy physics has been outlined here. We will now return to the strong nuclear force in more detail as this thesis will explore phenomena thereof.

2.2 Quantum Chromodynamics

In the framework of the QFT and the SM, the strong nuclear force is described by QCD [1]. Natural units will be used throughout this work; the numeric value of the speed of light c and the Planck constant $\hbar$ is set equal to one. The theory of the strong interaction is invariant under transformations of the $SU(3)$ gauge group. The Lagrangian can be written as

$$\mathcal{L}_{\mathrm{QCD}} = \bar{\psi}_i \left(i \left(\gamma^\mu D_\mu\right)_{ij} - m\delta_{ij} \right) \psi_j - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a \tag{2.2}$$

Where the quarks are represented by the spinor field ψ, along with their corresponding mass m. The so-called gamma or Dirac matrices are represented by γ^μ, they are essential for the description of spinor fields. The indices μ and ν indicate spacetime dimensions, while the summation over the indices i and j is a summation of the elements of a fundamental representation of the $SU(3)$ group. A fundamental representation is an irreducible finite-dimensional representation of a Lie group. Each element in a Lie group can be produced by the repeated action of infinitesimal elements, called the generators of the Lie group. In the case of the $SU(3)$ group, the generators have a non-zero commutator. Such Lie groups and QFTs are labeled as non-Abelian. An example of an Abelian QFT is Quantum Electrodynamics (QED), which describes the electromagnetic force and which is invariant under transformations of the $U(1)$ group. The partial derivative has been replaced by the covariant derivative D_μ in order to preserve the local gauge invariance and is defined as

$$D_\mu = \partial_\mu - ig\mathcal{A}^a_\mu t^a \tag{2.3}$$

Where the covariant derivative is defined in terms of the representation matrices t^a, while the gluonic field tensor takes the following form

$$G^a_{\mu\nu} = \partial_\mu \mathcal{A}^a_\nu - \partial_\nu \mathcal{A}^a_\mu + g f^{abc} \mathcal{A}^b_\mu \mathcal{A}^c_\nu \tag{2.4}$$

The gluonic fields are indicated by $\mathcal{A}^a_\mu$. In the case of an Abelian theory the third term would be absent, as the structure constants, represented by f^{abc}, that define the commutation relations between the generators would all be equal to zero. The gluonic fields are in the adjoint representation of the $SU(3)$ group, indicated by the indices a, b, c, … It is the irreducible representation to which the generators belong, therefore, the representation matrices are given by the structure constants themselves. For further details on Lie groups, Lie algebras and representations in the framework of QFT, we refer too [1,2].

A direct consequence of the non-Abelian character of QCD and the product of the two gluonic field tensors in eq. (2.2), is the self-interaction of the gluons. All interaction vertices of QCD are schematically depicted in Fig. 2.3. Due to the self-interaction of the gluons, the range of the strong nuclear force is smaller compared to the electroweak force, typically the size of the nucleus of an atom ($\sim 10^{-15}$m). Resulting in the prevention of colored particles from spreading out into the space around them. The anti-screening brought forth by the self-interaction is referred to as *color confinement* or just *confinement*, and is one of the two main properties exhibited by QCD. The second being *asymptotic freedom*. The strong nuclear interaction, just like all interactions, is characterized by a coupling constant α_{S}. The strong coupling constant varies as a function of the energy scale characterizing the interaction, often referred to as the renormalization scale μ_{R}. Their lowest order relation is expressed below.

$$\alpha_{\mathrm{S}} = \frac{12\pi}{(33 - 2N_{\mathrm{f}}) \ln \frac{\mu_{\mathrm{R}}^2}{\Lambda_{\mathrm{QCD}}^2}} \tag{2.5}$$

Where the parameter Λ_{QCD} is the QCD scale and N_{f} represents the number of quark flavours. A more precise expression can be obtained through the inclusion of higher-order corrections [55]. At high energies, the coupling constant decreases and the quarks and gluons start behaving like quasi-free particles, which is often referred to as *asymptotic freedom* and was discovered in 1973 by David Gross and Frank Wilczek [56], and independently by David Politzer [57]. In electromagnetism, the exact opposite behavior is observed, as the coupling constant α increases when the transferred momentum increases.

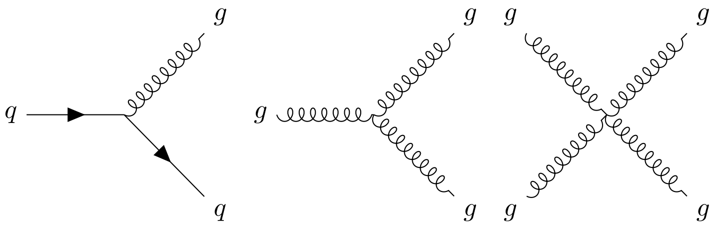


Figure 2.3: The three interaction vertices of QCD. From left to right: the quark-gluon, the three-gluon and four-gluon interaction vertex.

2.2.1 Quark Parton Model and Scaling Violations

During the 1960s, the role of QCD started to grow, as higher and higher energies were achieved at experiments, leading to the discovery that the proton is a composite object rather than a point-like one. A first attempt at understanding the inner structure of the proton was made by Feynman, Bjorken and Pachos in 1969 when they formulated the Quark Parton Model (QPM) following observations of electron-proton collisions at the SLAC experiment [58]. In the QPM, the proton is suggested to be composed of three bound quarks. In an electron-proton collision, one of the quarks would interact with the incoming electron, instead of the whole proton. The description of the collisions uses the structure function $F_2(x_{Bj}, Q^2)$, defined below.

$$F_2(x_{Bj}, Q^2) = x_{Bj} \sum_q e_q^2 f_q(x_{Bj}, Q^2) \tag{2.6}$$

Where x_{Bj} is the so-called x-Bjorken, representing the fraction of the momentum of the proton that the quark partaking in the interaction carries. While Q^2 is defined as the negative square of the four-momentum exchanged between the electron and the proton. The fractional quark charge is denoted as e_q. The function $f_q(x_{Bj}, Q^2)$ is the Parton Distribution Function (PDF) of the quarks. The PDF expresses the probability that a parton of the flavour q is involved in the interaction at the scale Q^2, while carrying a fraction x_{Bj} of the four-momentum of the proton from which it originates.

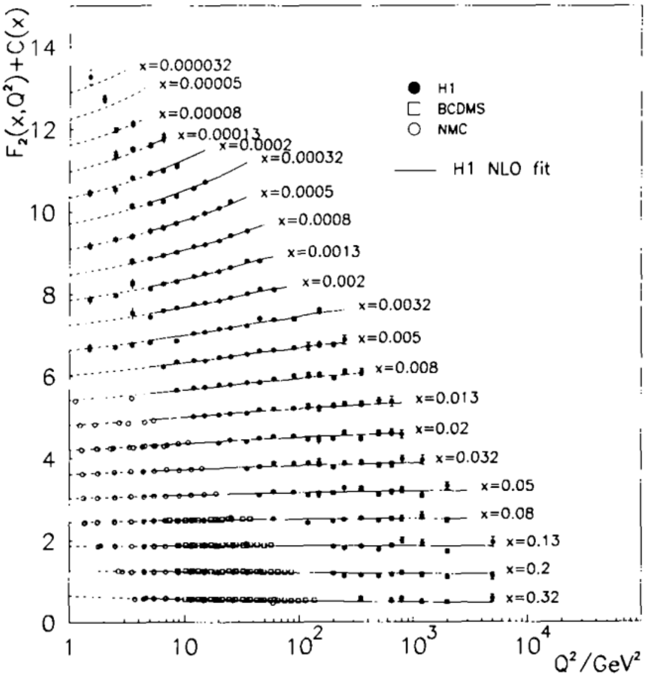


Figure 2.4: The results of a precision measurement of the structure function F_2 as a function of x_{Bj} and Q^2, taken from [59]. Scaling violations are observed as the structure function rises as a function of Q^2 as x_{Bj} becomes smaller.

In the QPM as described above, the structure function is expected be independent of the scale of the interactions Q^2. However, precision measurements were performed over many orders of magnitude of x_{Bj} and Q^2. The structure function was found to be scale dependent [59]. Measurements of the scale dependent structure function for different values of x_{Bj} and Q^2 are shown in Fig. 2.4. The scale dependence can be explained by the inclusion of the gluon in the QPM. The gluon binds the different quarks together, leading to a more fluent picture of the proton structure: a quark can emit a gluon at any point inside the proton, which in turn can split into a quark-antiquark pair. These quarks can interact with other quarks and gluons inside the proton and so on. However, the net amount of quarks remains equal to three, these three quarks determine the properties of the proton and are labeled as valence quarks. The additional quark-antiquark pairs and gluons are referred to as sea quarks and gluons, respectively. In general, the structure of the proton is probed more precisely at higher energies, resulting in the more complex picture described above and schematically depicted in the left plot in Fig. 2.5. The right plot shows the measurement of PDF as a function of x performed by the HERAPDF collaboration [60]. At large values of x, the valence quarks u_v and d_v dominate, while at lower values the gluon and the sea quark PDF start to contribute significantly.

When including the composite structure of the proton and the additional interactions to the lowest order electron-proton interaction, the scaling violation was naturally ob-

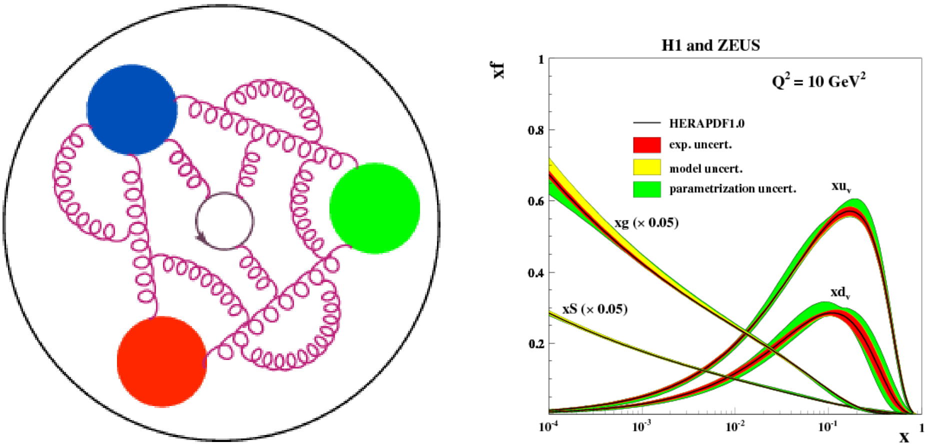


Figure 2.5: Left: a schematic depiction of the composite structure of the proton, taken from [61]. The red, blue and green valence quarks determine the properties of the proton. At high energies, the composite structure of the proton is probed further, revealing the presence of sea quarks and gluons. Right: PDF obtained by the HERAPDF collaboration at a scale $Q^2 = 10$ GeV^2, taken from [60]

tained. In Fig. 2.6, the electron-proton collision is shown without any corrections on the top left. The diagram is the Leading Order (LO) contribution, while all other diagrams are first order corrections, in other words: the diagrams are at Next-to-Leading Order (NLO). The possible emission of quarks and gluons by any parton is handled by the so-called splitting functions. However, one wants to include multiple emissions analytically, which is achieved by sets of evolution equations. Three different sets of equations will be discussed below. Each set is valid in a different kinematic region of the phase space, schematically shown in Fig. 2.7. The sets of evolution equations are obtained through a perturbative expansions in the strong coupling constant α_S, resulting in terms dependent on the momentum fraction x and the scale Q^2. A resummation of the terms is performed in three distinct manners, yielding the three sets of evolution equations. The approach to the resummation in all three cases is the same. The Leading Log Approximation (LLA) is used, so that the resummation only includes single logarithmic terms. To finish the discussion on evolution, an alternative to the framework of evolution equations will be discussed.

The DGLAP Evolution Equations

The Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations [62,63] are obtained by performing a resummation of terms of the type $(\alpha_S \ln (Q^2/Q_0^2))^m$ in the LLA, where the evolution starts at the scale Q_0^2 and ends at Q^2. Therefore, the evolution equations should be applied in a regime where $Q^2/Q_0^2 \gg 1$ and x_{Bj} is large enough to justify the neglect of terms proportional to $\ln(1/x_{Bj})$. The validity region of the DGLAP evolution equations is sketched in Fig. 2.7. The DGLAP equations can be written as

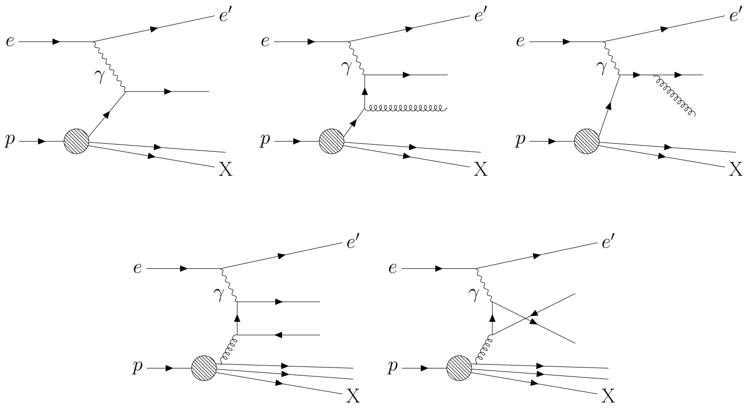


Figure 2.6: All possible contributions for electron-proton scattering, where the top left diagram is the LO contribution, while all other diagrams are at NLO. The electron before and after the emission of the photon γ is indicated with e and e', respectively. The proton before the interaction is marked as p, while the proton remnant is indicated with X. On the top row from left to right: a quark coming from the proton interacts with the electron, a quark emits a gluon prior to interacting with the electron, and the quark emits a gluon after interacting with the electron. The bottom left and right diagrams show the interaction of the photon with a sea quark.

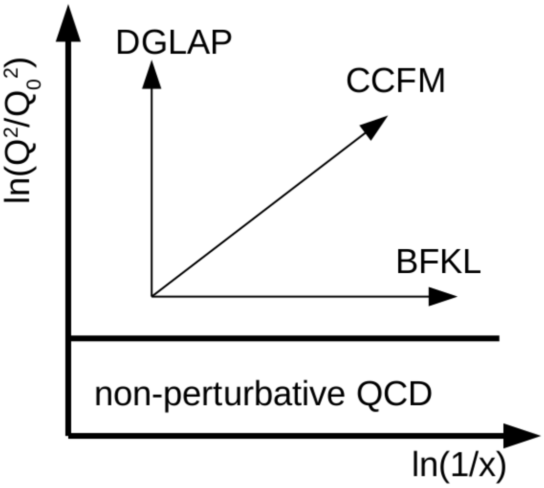


Figure 2.7: The regions where the three sets of evolution equations are valid is schematically depicted.

$$\frac{d}{dt}f_i(x,t) = \frac{\alpha_\mathrm{S}(Q^2)}{2\pi}\sum_j \int_x^1 \frac{dx'}{x'}f_i(x,t)P_{ij}\left(\frac{x}{x'}\right), \tag{2.7}$$

$$\text{where} \quad t = \ln{(Q^2/Q_0^2)} \tag{2.8}$$

Where x/x' is the momentum fraction of the emitted parton with respect to the mother parton, often denoted as z. The emission of the partons is governed by the splitting functions $P_{ij}(z)$, representing the probability of a mother parton i radiating a parton of the kind j with the momentum fraction z. The splitting functions can be written as a perturbative expansion in α_S.

$$P_{ij}(z) = P_{ij}^0(z) + \frac{\alpha_\mathrm{S}}{2\pi}P_{ij}^1(z) + \frac{\alpha_\mathrm{S}^2}{(2\pi)^2}P_{ij}^2(z) + \ldots \tag{2.9}$$

The splitting functions have been calculated up to the order α_S^3 [64]. Often in the LLA approximation, only the lowest order terms of the perturbative expansion are considered. The four splitting functions at LO are given below.

$$P_{gg}^0(z) = 6\left[\frac{z}{1-z} + \frac{1-z}{z} + z(1-z)\right] \tag{2.10}$$

$$P_{qg}^0(z) = \frac{1}{2}\left[z^2 + (1-z)^2\right] \tag{2.11}$$

$$P_{gq}^0(z) = \frac{4}{3}\frac{1+(1-z)^2}{z} \tag{2.12}$$

$$P_{qq}^0(z) = \frac{4}{3}\frac{1+z^2}{1-z} \tag{2.13}$$

Two different divergences are present in the splitting functions. The first divergency occurs for $z \to 1$, and is both a collinear and a soft divergence. On the one hand, the collinear singularity is cancelled through the inclusion of virtual corrections or loop diagrams. On the other hand, the soft divergence correspond to the emission of a soft parton and is treated by introducing the so-called plus prescription to the splitting functions.

$$P_{gg}^0(z) = 6\left[\frac{z}{(1-z)_+} + \frac{1-z}{z} + z(1-z)\right] \tag{2.14}$$

$$P_{qq}^0(z) = \frac{4}{3}\frac{1+z^2}{(1-z)_+} \tag{2.15}$$

$$\text{where} \quad \int_0^1 dz\frac{f(z)}{(1-z)_+} = \int_0^1 dz\frac{f(z)-f(1)}{1-z} \tag{2.16}$$

The need for the introduction of a soft cutoff is eliminated with the plus prescription. The treatment of the divergence due to a soft emission can be generalized through the use of the Sudakov form factor Δ_s [65], defined below.

$$\Delta_{\mathrm{s}}(Q^2, Q_0^2) = \exp\left(-\int_{Q_0^2}^{Q^2} \frac{dq^2}{q^2} \int_0^{z_{\max}} dz \frac{\alpha_{\mathrm{S}}}{2\pi} P(z)\right) \tag{2.17}$$

The Sudakov form factor represents the probability of no emission when a parton evolves from one scale Q_0^2 to another scale Q^2. The Sudakov form factor takes all virtual corrections of real emissions due to the divergence for $z \to 1$ into account and allows for the DGLAP equations to be rewritten into the integral form.

$$f(x, Q^2) = f_i(x, Q_0^2)\Delta_{\mathrm{s}}(Q^2, Q_0^2) + \int \frac{dq^2}{q^2} \int \frac{dz}{z} \frac{\Delta_{\mathrm{s}}(Q^2, Q_0^2)}{\Delta_{\mathrm{s}}(q^2, Q_0^2)} P(z) f\left(\frac{x}{z}, q^2\right) \tag{2.18}$$

In the part of the phase space where $z \to 0$, or the low-x regime, a second divergence arises, as the resummation of the terms of the type $(\alpha_{\mathrm{S}} \ln{(Q^2/Q_0^2)})^m$ in the LLA is no longer valid. Dedicated evolution equations have to be formulated to resolve the second divergence, which will be discussed in the sections to follow.

Due to the resummation in the LLA of the terms of the type $(\alpha_{\mathrm{S}} \ln{(Q^2/Q_0^2)})^m$, a strong ordering of the virtuality $k_{t,i}$ carried by the partons is obtained. An example of multiple emissions of gluons by a mother gluon, often referred to as a gluon ladder, is shown in Fig. 2.8. The strong ordering would mean that

$$|k_{t,0}^2| \ll |k_{t,1}^2| \ll \ldots \ll |k_{t,n-1}^2| \ll |k_{t,n^2}| \ll Q^2 \tag{2.19}$$

By evolving a parton from an initial scale Q_0^2 to Q^2, each successive emitted parton will carry a lower transverse momentum. Due to momentum conservation a second ordering, although weaker compared to the ordering on the virtuality, is present. The momentum fraction of the emitted partons decreases as well

$$x_0 > x_1 > \ldots > x_{n-1} > x_n \tag{2.20}$$

A important assumption in the derivation the DGLAP evolution equations is that the partons are assumed to travel in the same direction of the incoming proton. The transverse momentum of the emitted partons can be neglected and together with the fact that the virtuality of all partons is much smaller as Q^2, the partons can be treated as massless, effectively simplifying the calculations.

The BFKL Evolution Equation

Where the LLA approximation is made in terms of $(\alpha_{\mathrm{S}} \ln{(Q^2/Q_0^2)}^m$ for the DGLAP equation, the Balitsky-Fadin-Kuraev-Lipatov (BFKL) evolution equation [66, 67] is obtained by performing the LLA in terms of $(\alpha_{\mathrm{S}} \ln{(1/x)})^m$. The validity region compared to the other sets of evolution equations is shown in Fig. 2.7. The result is the presence of a strong ordering of the momentum fractions of the emitted partons. For the gluon ladder shown in Fig. 2.8 the ordering would entail that

$$|x_0| \gg |x_1| \gg \ldots \gg |x_{n-1}| \gg |x_n \tag{2.21}$$

Due to the different LLA, a strong ordering of the momentum fraction has effectively replaced the ordering of the transverse momentum of the emitted partons, obtained in the DGLAP equation. The BFKL equation is expected to be valid in the low-x regime, where $\ln(Q^2/Q_0^2) < \ln(1/x)$. As the DGLAP evolution breaks down in the low-x region of the phase space, the BFKL offers a better description of the evolution of the partons.

Since the virtuality of the partons does not have to follow the same ordering as in the DGLAP evolution equations, the virtuality and the transverse momenta can no longer be neglected. The PDF must have a dependence on transverse momentum of the partons. The Transverse Momentum Dependent Parton Distribution Function (TMD) is an extension of the PDF and resolves this issue, as the partons are treated off-shell. The TMD relates to the PDF for the gluon in the following way

$$xf_g(x, Q^2) = \int \frac{d^2k_t}{\pi} xA(x, k_t^2, Q^2)\Theta(Q^2 - k_t^2) \tag{2.22}$$

Where $A(x, k_t^2, Q^2)$ is the TMD for the gluon and Θ is the Heaviside step function. The BFKL evolution equation can be written as

$$\frac{d}{d\ln(1/x)} A(x, k_t^2, Q^2) = \int_0^\infty d^2k'_t A(x, k_t'^2, Q^2) K(k_t^2, k_t'^2) \tag{2.23}$$

The function $K(k_t^2, k_t'^2)$ is the BFKL splitting kernel. Only the gluon is considered, as the gluon density becomes dominant in the low-x regime compared to the quark distributions, as shown on the right of Fig. 2.5. When going to higher centre-of-mass energies in collisions, smaller values of x can be probed, resulting in a denser proton content. Eventually saturation effects [68] are expected to occur in order to save the unitarity, i.e. the interaction probability of two partons should always be smaller than one. Whenever the proton reaches a critical density, it is expected that the partons will start recombining. In calculations, the saturation of the gluon density is included by limiting the gluon TMD. Apart from saturation effects, the increased gluon density is expected to result in a larger probability for MPI. Therefore, the low-x regime is expected to be more sensitive to DPS (and MPI) effects.

The CCFM Evolution Equations

The Ciafaloni-Catani-Fiorani-Marchesini (CCFM) evolution equation [69, 70] forms the bridge between the two former sets of evolution equations. A resummation of leading log terms of the type $(\alpha_S \ln(Q^2/Q_0^2))^m$ and $(\alpha_S \ln(1/x))^n$ is performed. Therefore, the CCFM evolution is valid at both large Q^2/Q_0^2 and small x, effectively forming the bridge between the DGLAP and the BFKL evolution, as schematically shown in Fig. 2.7. The ordering of the emitted partons is not in $|k_t^2|$, nor in x, but rather in the emission angle. For the gluon ladder in Fig. 2.8, such an angular ordering would yield that

$$\xi_0 \ll \xi_1 \ll \ldots \ll \xi_{n-1} \ll \xi_n \ll \bar{\xi} \tag{2.24}$$

Where the angles $\xi_i = 1 - \cos\theta_i$, with θ_i the emission angle of the i'th gluon, and $\bar{\xi}$ the maximum allowed value determined by the interaction. It can be shown that the angular ordering condition above implies the ordering below.

$$q_0 z_0 < z_1, q_1 z_1 < z_2, \ldots, q_{n-1} z_{n-1} < q_n \tag{2.25}$$

The four momenta q_i are defined as $q_i^2 = -Q_i^2$ and Q_i^2 the scale of the i'th emission. The ordering is expressed by the Heaviside step function Θ and enters the CCFM equation, which can be formulated as

$$\begin{aligned} xA(x, k_t^2, Q^2) &= xA_0(x, k_t^2, Q^2)\Delta_{\mathrm{s}}(q) \\ &+ \int dz \int \frac{d^2q}{\pi q^2} \Theta(Q - zq)\Delta_s(Q, zq)\tilde{P}(z, q, k_t)A\left(\frac{x}{z}, k_t'^2, q^2\right) \end{aligned} \tag{2.26}$$

Where Δ_{s} is the aforementioned Sudakov form factor. The function $\tilde{P}(z, q, k_t)$ is the CCFM splitting function, defined below, while all other definitions remain as before.

$$\tilde{P}(z, q, k_t) = \frac{\bar{\alpha}_{\mathrm{S}}\left(q^2(1-z)^2\right)}{1-z} + \frac{\bar{\alpha}_{\mathrm{S}}(k_t^2)}{z}\Delta_{\mathrm{ns}}(z, q, k_t) \tag{2.27}$$

The running of the coupling is taken into account through $\bar{\alpha}_{\mathrm{S}} = 3\alpha_{\mathrm{S}}/\pi$, which takes the scales $q^2(1-z)^2$ and k_t^2 as input. The function Δ_{ns} is the so-called non-Sudakov form factor. Whereas the Sudakov form factor took all virtual corrections for the real emission for the divergence $z \to 1$ into account, the non-Sudakov form factor does the same but for the $z \to 0$ divergence.

The Color Dipole Model

Originally the Color Dipole Model (CDM) [71–73] was formulated to provide an alternative description for the evolution of partons for $e^+e^- \to q\bar{q}$ processes. The CDM allows for the calculation of higher order corrections of QCD emissions, while not being based on the framework of the evolution equations presented above. A color dipole is spanned between the initial quark-antiquark pair, which is linked to a gluon emission probability. Whenever a gluon is emitted, two new dipoles are formed between each of the initial partons and the emitted gluon. The process is repeated until an avalanche or cascade of particles is formed, where each emitted particle is softer compared to the previous one. A visual representation of the workings of the CDM is depicted in Fig. 2.9. Later the framework was extended to electron-proton and proton-proton collisions.

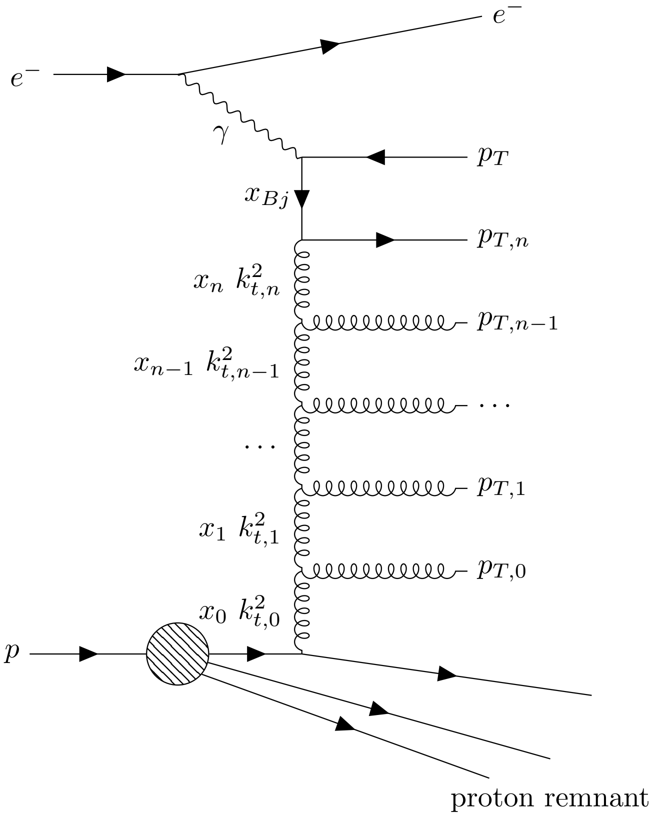


Figure 2.8: A visual representation of an electron-proton interaction along with a gluon ladder. A gluon coming from the proton emits multiple gluons with different transverse momenta in succession before interacting with an electron through the exchange of a photon. Each emission subsequently changes the transverse momentum and the momentum fraction of the mother parton. The virtuality of each parton is indicated with k_i^2 along with its momentum fraction x_i, while the transverse momentum of the emitted partons is indicated with $p_{\mathrm{T,i}}$.

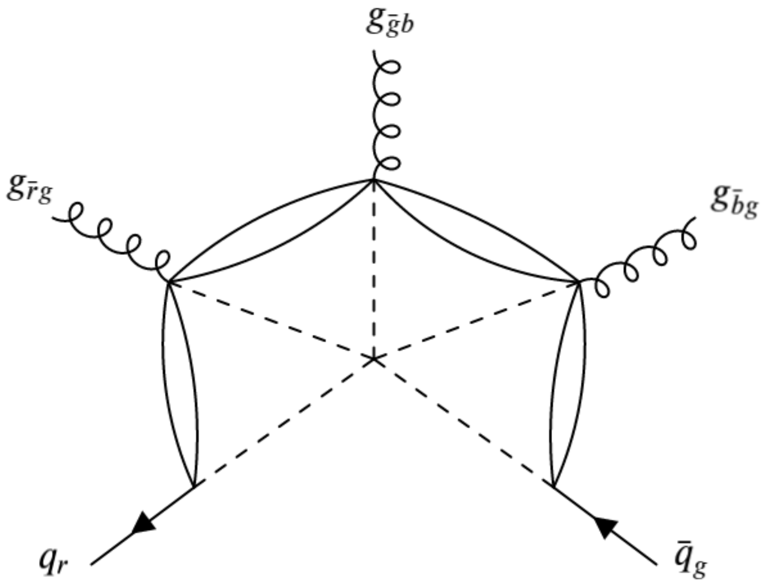


Figure 2.9: A visual representation of the workings of the CDM. The initial quark-antiquark pair spanning the first color dipole is shown on the bottom of the figure. After each gluon emission, new color dipoles are spanned between the particles. A possible example of the color charges of the particles is given in the figure. The CDM provides an alternative approach to higher order particle emissions and particle evolution.

Chapter 3

Hadron Collisions

In the previous chapter, the underlying theory was discussed. However, the whole framework in which measurements are performed and predictions are made still needs to be highlighted. In this chapter, we will start by defining factorization, which effectively allows a hadronic collision to be split into multiple parts that can either be measured and modeled from experiments or can be calculated perturbatively. Next, we will introduce the DPS formalism along with the observables chosen for this analysis. Afterward, all models used to make DPS predictions will be introduced. Lastly, a generator level study of the DPS-sensitive observable is performed.

3.1 Factorization

The factorization theorem [74] states that the cross section of a process can be separated into two parts, the short-distance partonic cross section and universal long-distance functions describing the other spectator partons. The cross section characterizing a hadron-hadron interaction can be written as the product of the perturbatively calculable partonic cross section and the non-perturbative PDF describing the hadron structure. Since the scale of the perturbative part is in general larger compared to the scale of the non-perturbative part, the former is referred to as the hard part of the interaction, while latter is often labeled as the soft part. The general form of the hadronic cross section of an interaction then takes the form of

$$\sigma_{\mathrm{h_1h_2}} = \sum_{i,j} \int dx_1 dx_2 f_{i/h_1}(x_1, \mu_F^2) f_{j/h_2}(x_2, \mu_F^2) \frac{d\hat{\sigma}_{ij\to k}}{dx_1 dx_2} \tag{3.1}$$

where $x_{1,2}$ is the longitudinal momentum fraction that the partons receive from their respective mother hadron. The cross section $\hat{\sigma}_{ij\to k}$ is the partonic cross section for the process where partons i, j form a given final partonic state k. Partonic cross sections are perturbatively calculable, they consist of the Matrix Element (ME) and a Lorentz-invariant phase space factor Φ_k. They are discussed in more detail in the next section. The functions f_{i/h_1} and f_{j/h_2} are the PDF, which give the probability that a parton i, j, coming from the hadron $h_{1,2}$, takes part in the hard interaction. The PDF takes the factorization scale as a second input, μ_F^2 characterizes the scale at which the separation between the hard and the soft part is made. It is often set equal to the scale of the hard interaction, but no fixed or standard recipe exists. Since the PDF describes the soft or non-perturbative part of the hadronic interaction, they can not be calculated and have to be extracted from data. Many different sets of PDF have been developed. The PDF sets used in this analysis will be listed when discussing the Monte Carlo (MC) models in Sec. 3.7.

In the formulation of the hadronic cross section in eq. (3.1), the transverse momentum of the partons in question has been neglected. This approach to factorization is referred to as collinear factorization. The approximation is motivated by the fact that at high-energy collisions the scale of the transversal momentum of the partons is much smaller compared to the energy scale of the hard interaction. Collinear factorization can, therefore, fail whenever the size of the transverse momentum vector of the partons becomes comparable to the size of its longitudinal momentum, i.e. the low-x regime, which will be probed in this analysis. The DGLAP evolution equations are used in the collinear factorization scheme as the transverse momentum of the partons is neglected. Different approaches to incorporate the transverse momentum exist, such as transverse momentum dependent factorization [74] or high-energy factorization (k_T-factorization) [75–77]. One has to switch from the standard PDF to the TMD, which generally relate to the standard PDF as

$$f_{i/h}(x_i, \mu_F^2) \sim \int d\vec{k}_\mathrm{T} f_{i/h}(x_i, \vec{k}_\mathrm{T}, \mu_F^2) \tag{3.2}$$

An explicit example was given in the context of the BFKL evolution equation, however, in the CCFM framework TMDs have to be used as well. The exact relation between the PDF and the TMD for the gluon is given in eq. (2.22).

Both collinear factorization and k_T-factorization, will be used to study four jet production in proton-proton collisions. The models will be detailed in Sec. 3.7

3.2 Hard Interaction

The partonic cross section in the formula of the cross section (3.1) can be written as

$$\hat{\sigma}_{ij\to 2+n} = \int d\Phi_{2+n} \left|\mathcal{M}_{ij\to 2+n}\right|^2 \tag{3.3}$$

Where $d\Phi_{2+n}$ is the Lorentz invariant differential phase space factor of the final state containing $2+n$ partons, which has been used instead k for reasons that will become clear further down in the text. The hadronic cross section can be calculated through perturbation theory where the ME $\mathcal{M}_{ij\to 2+n}$ can be written as the sum over all possible Feynman diagrams or

$$\hat{\sigma}_{ij\to 2+n} = \int d\Phi_{2+n} \left|\sum_l \mathcal{F}^{(l)}_{ij\to 2+n}\right|^2 \tag{3.4}$$

Where $\mathcal{F}^{(l)}_{ij\to 2+n}$ is a Feynman diagram with $2+n$ partons in the final state and with l additional loops. A loop is called a virtual correction since it does not add any real partons to the final state. The number n can be seen as the number of additional legs to the $2 \to 2$ ME which corresponds to the number of real corrections added to the $2 \to 2$ ME. When certain values for the numbers n and l are chosen, different processes are included in the analytical calculation of the ME. It is often stated that the calculation goes up to a fixed order in α_S, which refers to the number of QCD vertices included in the calculation of the Feynman diagrams. Each virtual correction increases the order in α_S with two units, while the real corrections add one unit to the order of α_S. In Fig. 3.1, a $2 \to 2$ ME is written as the perturbative series over all Feynman diagrams with an increasing number of virtual corrections or loops l.

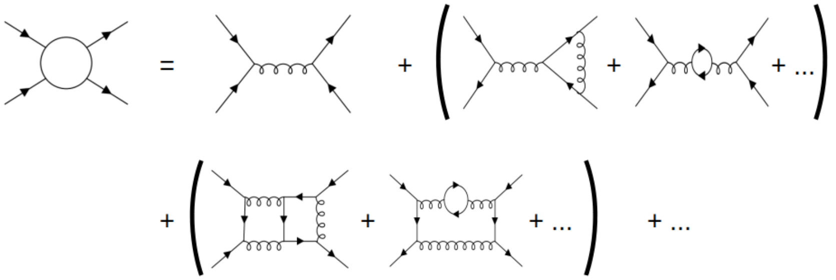


Figure 3.1: A $2 \to 2$ ME written as the perturbative series over all Feynman diagrams. The ME from eq. (3.3) with $n = 0$, i.e. $\mathcal{M}_{ij\to 2}$, is shown on the top left side and is written as a perturbative series over all $2 \to 2$ Feynman diagrams. The first term is the LO diagram for which $l = 0$, while the second and third terms, the two groups of diagrams between the parentheses, include the one- and two-loop corrections for which $l = 1$ and $l = 2$ respectively. The second and third term represent the NLO and N^{2}LO corrections.

Three general classes depending on the choice of n and l can be defined and are listed below.

- $n, l = 0$: production of the final state with 2 partons at LO. Only the $2 \to 2$ diagram at LO is included in the calculation, which corresponds to the LO diagram in Fig. 3.1.
- $n \neq 0$, $l = 0$: production of a final state with $2+n$ partons at LO. A number of n real corrections to the LO $2 \to 2$ diagram is included in the analytical calculation of the ME.
- $n, l \neq 0$: production of a final state with $2+n$ partons at N^{l}LO which also includes the production of a final state with $2+n+1$ partons at N^{l-1}LO, and so on until the production of a final state with $2+n+l$ at LO.

In this work, MC event generators at LO and NLO will be used in combination with $2 \to 2$, $2 \to 3$ and $2 \to 4$ ME. All models will be discussed in detail in Sec. 3.7

3.2.1 Parton Showers

To incorporate or to simulate the effect of the higher-order diagrams, parton shower algorithms [78], are used. The showering algorithms are based on the different types of parton evolution and evolve the initial and final state partons in general differently, thus one distinguishes between Initial State Radiation (ISR) and Final State Radiation (FSR).

The FSR is described by determining the maximum scale Q^2, which is obtained from the hard scattering, and letting it evolve down to a lower cutoff scale Q_0^2. In MC generators the evolution is performed by examining the probability that a parton does not emit another parton while going one scale Q_1^2 to another scale Q_2^2, the probability can be written as

$$P_i(Q_1^2, Q_2^2) = f_{i/h}(x_i, Q_2^2) \exp\left(- \int_{Q_1^2}^{Q_2^2} f_i(t) \right) dt \tag{3.5}$$

The exponent is the aforementioned Sudakov form factor, which represents the probability of a parton i evolving from the initial scale Q_1^2 to another scale Q_2^2 without the occurrence of a branching, i.e. without the emission of a parton.

For the description of the ISR, one uses a backward evolution. First the hard scattering is selected, after which the parton shower that came before the scattering is reconstructed. The reconstruction starts at the hard scale of the scattering Q^2 and evolves backwards to the initial softer cutoff scale Q_0^2. The cutoff scale regulates the soft and collinear divergences.

3.2.2 Matching Procedures

Since the analytical calculation of all order diagrams is not possible, the parton showers are used to simulate the contribution from the missing diagrams. However, one cannot match the ME with the parton shower in a naive manner since double counting can occur. The matching procedure, to which different approaches exist, aims to resolve the double counting. The so-called MLM matching scheme [78, 79] will be used to match LO $2 \to 2 + n$ ME with the partons shower. The method is based on an event rejection method. Whenever a parton is emitted by the parton shower above the scale of the partons in the ME, the event is discarded. Additional approaches have been developed to perform the matching at NLO and will be used in this work, i.e. the POWHEG [80–82] and `MCNLO` [83, 84] methods. Both will be highlighted further when discussing the Monte Carlo event generators, used to make predictions in Sec. 3.7.

3.3 Underlying Event

The Underlying Event (UE) is used to indicate a collection of phenomena apart from the initial hard interaction, occurring in hadronic collisions. These phenomena are partonic interactions, which underlie the hard interaction, hence the name. Generally, the partonic interactions constituting the UE take place at a lower scale compared to the initial hard interaction. Parts of the UE still fall into the perturbative framework, while other phenomena need to be parametrized as their typical scales are too soft. The distinction between three types of phenomena is made: the aforementioned ISR and FSR, the Beam-Beam Remnants (BBR) and Multiple Partonic Interactions (MPI). In Fig. 3.2, a sketch of a complete hadronic interaction with the three groups of phenomena is shown, all groups will be discussed below.

The ISR and FSR describe the emittance of quarks and gluons by the partons that take part in the hard interaction, as described in Sec. 3.2.1. When examining final states with more than two objects, the role of ISR and FSR becomes more important. The parton shower can add real emissions to the incoming or outgoing particles of the ME, effectively incorporating the contribution of higher-order diagrams to the ME. For example, the role of the ISR and FSR can not be neglected in 2 → 2 processes since at least two jets need to originate from the PS in order to obtain a four jet final state. Generally, the scale of the jets produced by ISR and FSR is lower compared to the scale of jets originating from the hard interaction.

Not all processes need both ISR and FSR, i.e. a process with a leptonic (hadronic) initial state and a hadronic (leptonic) final state will only be susceptible to FSR (ISR). Therefore, they can be studied independently: ISR can be studied through Drell-Yan production[1] in proton-proton collisions, while quark production in leptonic collisions allows for the study of FSR effects. In Fig. 3.2, the ISR and FSR are depicted by the blue lines, entering and leaving the hard interaction which is shown in red.

When protons interact, not all partons partake in the hard interaction, effectively leaving colored groups of partons or the BBR behind. The remaining partons are often referred to as spectator partons and can interact among themselves which needs to be modeled. As the BBR move predominantly in the direction of the initial incoming particle beams, its effects can only be measured by detector parts close to the beamline. Additionally, the typical scale of the BBR is much smaller compared to the hard interaction, therefore, the BBR mainly contributes to the soft part of the UE. In Fig. 3.2, the BBR is depicted by the blue lines going to the center of the picture which originate from the incoming partons without interacting with other partons.

A second possibility for the spectator hadrons is that they initiate additional hard interactions inside the same hadronic collision which is described by MPI. The colliding protons are composite objects containing partons of which the density can become large, e.g. the low-x regime where the gluon density grows exponentially, increasing the probability for the occurrence of MPI. The scale of MPI is smaller compared to the scale of the initial hard interaction. However, the scale of MPI can range from roughly the scale of the hard interaction all the way down to very low scales, effectively contributing to the hard and soft parts of the UE. In Fig. 3.2, MPI are shown in purple, on the bottom of the sketch.

From an experimental point of view, the existence of MPI has been presupposed as well. The total cross section of the hard interaction σ_{hard} is related to the generic partonic cross section $\hat{\sigma}$, introduced in eq. (3.1) as given below.

$$\sigma_{\text{hard}}(p_{\text{T,min}}) \sim \int_{p^2_{\text{T,min}}}^{s/4} \frac{d\hat{\sigma}}{dp^2_{\text{T}}} dp^2_{\text{T}} \sim \frac{1}{p^2_{\text{T,min}}} \tag{3.6}$$

---

[1] In a Drell-Yan process, a quark-antiquark pair produces a virtual photon or Z boson which decays into a lepton pair, e.g. $q\bar{q} \to Z \to e^+e^-$.

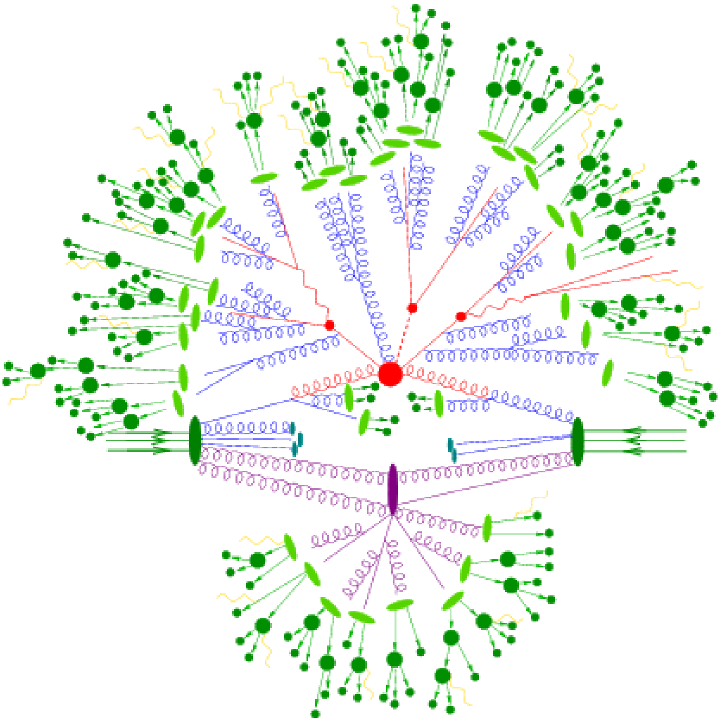


Figure 3.2: A sketch of a hadron collision with all its components, taken from [85]. Two partons coming from the inbound hadrons partake in the hard interaction (red). The hadrons emit partons before and after the hard interaction in processes labeled as ISR and FSR, both depicted in blue around the hard interaction. The BBR are the blue lines coming from the hadrons which do not interact further, while the purple lines represent MPI. The transition from the light green blobs to the dark green cascades represents the hadronization, where colored particles are converted in color neutral hadrons.

Where $p_{\mathrm{T,min}}$ is the minimal exchanged transverse momentum between the two partons, introduced as a cutoff in order to regulate the total partonic cross section of the hard interaction σ_{hard}. While the integrals over the momentum fractions $x_{1,2}$ in eq. (3.1) have been replaced by an integral over the transverse momentum squared. At values of $p_{\mathrm{T,min}} \approx 2-5$ GeV, the cross section σ_{hard} is found to be larger than the total inelastic cross section σ_{inel} which has been measured to be around 70-80 mb at a centre-of-mass energy of $\sqrt{s} = 13$ TeV [86–88]. However, this observation does not lead to contradictions since σ_{hard} is a partonic cross section rather than the hadronic cross section. Hadrons can be considered as bundles of partons, with a non-zero probability of MPI occurring simultaneously in a single collision. In this framework, the ratio of both cross sections can be larger than unity and can be viewed as the average number of parton interactions as a function of p_{T}:

$$\bar{n} = \frac{\sigma_{\mathrm{hard}}(p_{\mathrm{T}})}{\sigma_{\mathrm{inel}}} \tag{3.7}$$

In the simplest models of MPI and DPS, the different interactions are assumed to be completely independent of each other. Possible correlations between the partons coming from the same hadron are ignored, see Sec. 3.6 for more details. The result of this assumption is that n follows a Poisson distribution. However, the divergent behavior of the cross section σ_{hard} still remains and is regulated by the p_{T} cutoff. From a physical point of view, the cutoff is motivated by the fact that at increasingly smaller p_{T} of the partons the corresponding transverse wavelength becomes longer, therefore, the partons can no longer distinguish between the different color charges, effectively decreasing the coupling. The effects of MPI are expected to become increasingly important as the centre-of-mass energy of collisions rises since the low-x region becomes more accessible, resulting in higher parton densities.

In order to model the phenomena described above, specific measurements are performed at different centre-of-mass energies since the UE activity and the role played by the UE event in collisions are expected to increase when the centre-of-mass energy is raised. Measurements of the UE are performed in different channels due to the scale dependency of the UE, such measurements at a centre-of-mass energy of $\sqrt{s} = 13$ TeV include the study of leading particles and jets [89], of Drell-Yan processes [90] and of $t\bar{t}$ pair production [91]. In turn, these measurements are used to model the UE. The MC event generators PYTHIA 8 [92, 93] and HERWIG [94, 95] have adjustable parameters to control the behavior of their event modeling. The sets parameters are altered to make the models fit some aspects of the data better, such a collection of parameters is referred to as a tune, see Sec. 3.7 for more details.

3.4 Hadronization

The hard interaction, together with the parton shower and the UE description, do not yet give the full description of hadronic interactions. Due to confinement, all formed partons need to be converted into colorless hadrons. The conversion is referred to as hadronization. In Fig. 3.2, the hadronization step is shown at the outer edges, where the light green is converted into the dark green, representing colorless hadrons. Hadronization is not perturbatively calculable and needs to be modeled as well. Two main classes of models exist, the so-called string and cluster models.

The former class of models is based on the Lund model [96], which describes confinement linearly. Quarks are imagined to be encapsulated by a color flux tube. As they move away from each other, the energy of the color flux tube will keep rising up until the point that it becomes energetically favorable to form a quark-antiquark pair, effectively breaking up the string into two shorter pieces. The process is repeated until stable hadrons can be formed. The fragmentation function, defined below, represents the probability of the splitting of a string where a hadron with momentum fraction z of the original string is created.

$$f(z) \sim \frac{(1-z)^a}{z} e^{-b m_{\mathrm{T}}^2 / z} \tag{3.8}$$

Where m_{T} is the transverse mass of the hadron, equal to $\sqrt{m_h^2 + p_{\mathrm{T}}^2}$. While a and b are tune-able parameters. Baryon production can be incorporated by allowing the formation of diquark pairs in the splitting of the strings.

The cluster model [97–100] is based on the so-called preconfinement property of parton shower [101]. To initiate the hadronization process, the cluster model starts by non-perturbatively splitting gluons, coming from the parton shower, into quark-antiquark pairs in a color singlet state. Next, the neighboring partons are combined into singlet clusters. These clustered partons then undergo isotropic decays into pairs of hadrons. Compared to the string models, the cluster model has fewer parameters, however, it has problems in dealing with the decay of more massive clusters.

Fig. 3.3, taken from [102], schematically depicts the working of the string models and the cluster models.

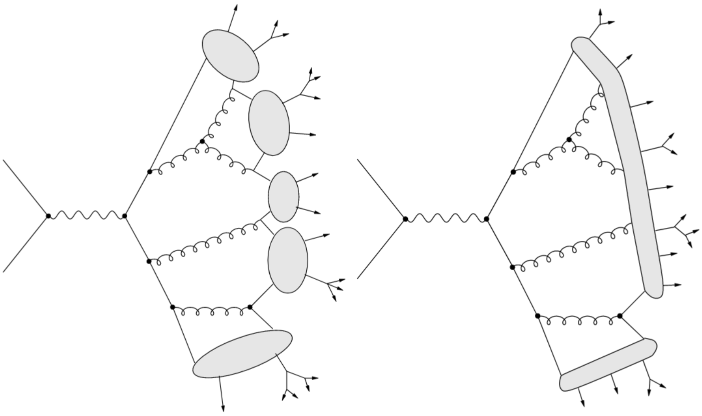


Figure 3.3: Schematic depiction of the differences between the string models (left) and the cluster models (right), taken from [102].

3.5 Defining a Jet

Once all partons from a collision have been hadronized, collections of hadrons in the vicinity of the original direction of the partons remain. From both an experimental and a theoretical point of view, it is advantageous to try and group these particles into a single object, matching the initially produced partons. In an experiment, neighbouring signals in the detector are recombined into one signal, while in theoretical calculations one can avoid working with a multitude of hadrons, effectively desensitizing the calculations from the hadronization process. The collections of particles, corresponding to an initially produced parton, are referred to as jets. Different types of jet algorithms have been developed over the years and are judged according to multiple criteria [103], given below.

- The infrared and collinear singularities should not appear in the perturbative calculations, as well as the algorithm should find solutions independent of infrared and collinear emissions.

- The solutions found by the algorithm should not depend on the frame of reference. If one were to boost the frame in the longitudinal direction, the same jet collections should be found.
- The quantities used to define the boundaries of the jets, should not be dependent on the exact final state. In this regard the total transversal energy would not be a proper quantity, as it depends on the exact amount of particles taken into account.
- The same jet collections for a certain collision should be found at all levels, i.e. parton, particle and detector level.
- A straightforward implementation of the algorithm is desired.

In this analysis, it is opted to use the anti-k_t algorithm [104], which is a sequential recombination jet algorithm that meets all of the criteria. In general, sequential recombination algorithms use a fixed measure in the recombination of two objects into a new object or jet. In regard to the clustering of the jets, the anti-k_t algorithm uses the distances defined below.

$$\begin{aligned} d_{ij} &= \min\left(p_{T,i}^{2p}, p_{T,j}^{2p}\right) \frac{\Delta_{ij}^2}{R^2} && (3.9) \\ d_{iB} &= p_{T,i}^{2p} && (3.10) \\ d_{\min} &= \min\left(d_{ij}, d_{iB}\right) && (3.11) \end{aligned}$$

Where $\Delta_{ij} = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ is the spatial distance between the partons i and j. The angle ϕ_i represents the azimuthal angle of the objects and R is a constant representing the jet cone radius. The parameter p is set equal to -1 for the anti-k_t algorithm. The distance measure of other sequential recombination algorithms differs only in their value of p. The k_t algorithm [105] uses $p = 1$, while the so-called Cambridge-Aachen algorithm [106] has implemented a value of $p = 0$. Lastly, y_i is a spatial coordinate defined as

$$y = \frac{1}{2} \ln\left(\frac{E + p_z}{E - p_z}\right) \tag{3.12}$$

The rapidity describes how forward/backward an object is with respect to the beam axis. In the limit where the mass of the particles is negligible, the rapidity can be approximated by the pseudorapidity which is defined below.

$$\eta = \frac{1}{2} \ln\left(\frac{\vec{p} + p_z}{\vec{p} - p_z}\right) = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \tag{3.13}$$

Where p_z is the momentum in along the beam axis and θ is the polar angle. The pseudorapidity itself is not invariant under Lorentz boosts along the z-axis, however, differences in the pseudorapidity are.

The anti-k_t algorithm uses the measure to calculate both distances for objects i and j. Whenever $d_{\min}$ is equal to d_{ij}, then they are combined into an object through the addition of the four-vectors of the momentum. If $d_{\min}$ is equal to d_{iB}, then the object i

is seen as a new jet and removed from the objects over which the algorithm is performed. The recombination procedure starts with the objects with the highest p_{T} and continues until objects cannot be combined any further. An example of the performance of the anti-k_t algorithm with a cone size $R = 1$ is given in Fig. 3.4. A more detailed discussion on the performance compared to other jet algorithms can be found in [104].

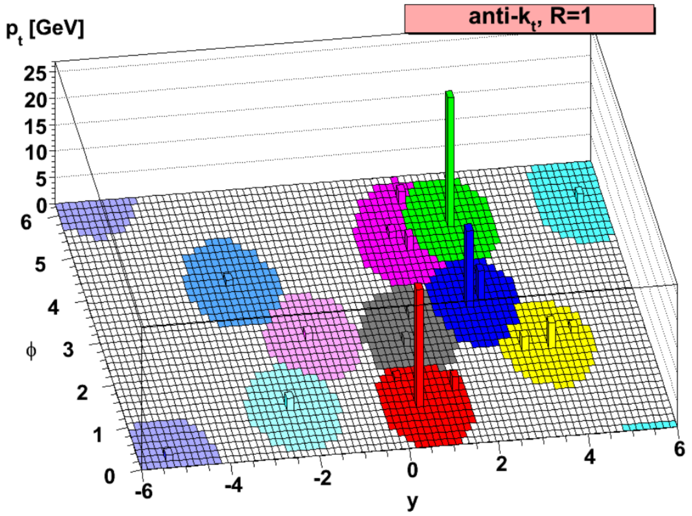


Figure 3.4: A parton level event shown in the $y - \phi$ space, where the jets have been clustered with the anti-k_t algorithm with a parameter $R = 1$, taken from [104]. As the algorithm starts with objects with the highest p_{T}, the jets associated with these initial objects tend to be circular. While the jets associated with the lower p_{T} objects were formed with the "leftovers".

3.6 Double Parton Scattering

With the most important aspects of hadron collisions defined and discussed above, an overview of the DPS formalism will discussed. In this section, an overview of the lowest order theory on DPS will be given. Afterward, earlier measurements and the DPS channel of choice will be highlighted. Lastly, multiple DPS-sensitive observables, used to study DPS in this work, will be defined.

3.6.1 The Double Parton Scattering Cross Section

In the framework of collinear factorization [74], the hadronic DPS cross section can be written as the product of two independent partonic cross sections and the so-called Double Parton Distribution Function (dPDF) [107], as given below.

$$\sigma^{\mathrm{DPS}}_{\mathrm{A,B}} = \frac{m}{2}\sum_{i,j,k,l}\int dx_1 dx_2 dx_1' dx_2' d^2\mathbf{b}\ \Gamma_{ij}(x_1, x_2, \mathbf{b}, \mu_F^{\mathrm{A}}, \mu_F^{\mathrm{B}})\Gamma_{kl}(x_1', x_2', \mathbf{b}, \mu_F^{\mathrm{A}}, \mu_F^{\mathrm{B}})$$

$$\hat{\sigma}^{\mathrm{A}}_{ik}(x_1, x_1', \mu_F^{\mathrm{A}})\hat{\sigma}^{\mathrm{B}}_{jl}(x_2, x_2', \mu_F^{\mathrm{B}}) \tag{3.14}$$

The $x_{1,2}$ and $x'_{1,2}$ are the longitudinal momentum fractions that the partons receive from their mother hadron, while b represents the impact parameter. The scales μ_F^A and μ_F^B characterize the processes A and B. They are their respective factorization scales. Both scales serve as the input for the aforementioned dPDF, which contains correlations between the two partons coming from each of the two hadrons. Both processes A and B are described by their own perturbatively calculable partonic cross sections $\hat{\sigma}^{\mathrm{A}}$ and $\hat{\sigma}^{\mathrm{B}}$. An additional symmetry factor m is introduced to prevent double counting, as the processes A and B can be identical. Whenever the two processes are distinguishable, the symmetry factor is equal to two and one otherwise.

Two ansatzes are made in order to rewrite the DPS cross section. The first being that the dependence on the transverse distance b factorizes with the rest of the dPDF in the following way

$$\Gamma_{ij}(x_1, x_2, \mathbf{b}, \mu_F^{\mathrm{A}}, \mu_F^{\mathrm{B}}) = F_{ij}(x_1, x_2, \mu_F^{\mathrm{A}}, \mu_F^{\mathrm{B}})G(\mathbf{b}) \tag{3.15}$$

where the function $G(b)$ is assumed to be universal and describes the dependence of the DPS cross section on the relative transverse distances between the partons. The second ansatz neglects the correlation between the two partons coming from the same hadron, allowing the function F to be formulated as the product of two PDF.

$$\Gamma_{ij}(x_1, x_2, b, \mu_F^{\mathrm{A}}, \mu_F^{\mathrm{B}}) = f_i(x_1, \mu_F^{\mathrm{A}})f_j(x_2, \mu_F^{\mathrm{B}})G(\mathbf{b}) \tag{3.16}$$

Applying these assumptions to the DPS cross section (3.14), results in the expression below.

$$\begin{aligned}\sigma^{\mathrm{DPS}}_{\mathrm{A,B}} &= \frac{m}{2}\sum_{i,j,k,l}\int dx_1 dx_2 dx_1' dx_2' d^2\mathbf{b} f_i(x_1, \mu_F^{\mathrm{A}})f_j(x_2, \mu_F^{\mathrm{B}})f_k(x_1', \mu_F^{\mathrm{A}})f_l(x_2', \mu_F^{\mathrm{B}})G^2(\mathbf{b}) \\ &\quad \times \hat{\sigma}^{\mathrm{A}}_{ik}(x_1, x_1', \mu_F^{\mathrm{A}})\hat{\sigma}^{\mathrm{B}}_{jl}(x_2, x_2', \mu_F^{\mathrm{B}}) \qquad (3.17) \\ &= \frac{m}{2}\int d^2\mathbf{b}\ G^2(\mathbf{b})\left(\sum_{ik}\int dx_1 dx_1' f_i(x_1, \mu_F^{\mathrm{A}})f_k(x_1', \mu_F^{\mathrm{A}})\hat{\sigma}^{\mathrm{A}}_{ik}(x_1, x_1', \mu_F^{\mathrm{A}})\right) \\ &\quad \times \left(\sum_{jl}\int dx_2 dx_2' f_j(x_2, \mu_F^{\mathrm{B}})f_l(x_2', \mu_F^{\mathrm{B}})\hat{\sigma}^{\mathrm{B}}_{jl}(x_2, x_2', \mu_F^{\mathrm{B}})\right) \qquad (3.18)\end{aligned}$$

One can identify the hadronic cross sections of the process A and B separately, i.e. the integral over the product of the partonic cross section together with the PDF for each of the two processes. After defining the effective cross section $\sigma_{\mathrm{eff}} = (\int d^2bG^2(b))^{-1}$, the so-called pocket formula of DPS is obtained.

$$\sigma_{\mathrm{A,B}}^{\mathrm{DPS}} = \frac{m}{2}\frac{\sigma_{\mathrm{A}}\sigma_{\mathrm{B}}}{\sigma_{\mathrm{eff}}} \tag{3.19}$$

Where the effective cross section σ_{eff} carries the information about the transverse distance between partons, however, it is obtained in a rather naive description of DPS. The validity of the DPS pocket formula and therefore the validity of the two ansatzes is discussed in [108]. It was found that the formula upholds and remains process independent only if inclusive scenarios are considered. Events cannot be vetoed due to the presence of additional objects in the final state. However, efforts to improve the description of DPS have been made, where the partons are not seen as completely uncorrelated. Although significant consequences of the incorporation of the correlations have been found [109–112], the models are still in an early stage.

The DPS cross section $\sigma_{\mathrm{A,B}}^{\mathrm{DPS}}$ and the effective cross section σ_{eff} have been determined for multiple final states at different centre-of-mass energies, using a range of different extraction techniques. Regardless of which is applied, the cross sections of the processes A and B in the DPS cross section need to be identified and measured. We will come back to the DPS pocket formula and the definition of the processes A and B when discussing the applied strategy for the extraction of the effective cross section σ_{eff} in Chap. 8.

3.6.2 DPS and σ_{eff} Measurements

Throughout the years, multiple measurements of DPS have been performed at varying centre-of-mass energies, and evidence for DPS events has been observed in multiple channels or final states. An overview of DPS measurements using multiple final states has been given in Chap. 1.

All measurements have provided DPS-sensitive distributions and have concluded that the addition of DPS to the simulation helps to improve the description of the data. Most of the studies have been able to extract a value for the parameter σ_{eff} in the framework of the DPS pocket formula, described above. Two main techniques have been applied in the extraction efforts: the template method and the tuning method. The former relies on the measurement of the differential cross section as a function of observables that are expected to be sensitive to DPS events in a certain region. Next, different background and signal templates have to be constructed. The background template is to simulate the SPS events, while the DPS processes are put in the signal template. The fraction of DPS events is determined by fitting the two templates to the differential cross section of the DPS-sensitive observable. Along with the measurement of the independent processes A and B, the effective cross section can be determined from the DPS pocket formula (3.19). While in the latter method varies parameters of the UE tune in order to make the predicted differential cross section as a function of the DPS-sensitive observable(s) fit the data better, effectively producing a new "DPS" tune. In this work, it is opted the template method and perhaps a more naive method that compares the shape of the predicted differential cross section as a function of DPS-sensitive observables with the data. Both will be discussed more extensively in Chap. 8.

Additional measurements that test the framework of the DPS pocket formula in different channels and at different energies are important from an experimental point of view. If the values of the parameter σ_{eff} appear to differ for different channels or at different centre-of-mass energies, it would indicate a clear need for improved DPS models.

This thesis aims to contribute to these previous measurements and to shed more light on DPS. Multiple differential cross sections as a function of a multitude of DPS-sensitive observables will be determined in the inclusive four jet channel. While the template method will be implemented for the extraction of the σ_{eff}.

3.6.3 DPS in Inclusive Four Jet Production

At the LHC, jet production is one of the most abundant processes and can be particularly relevant for DPS. Events with jets with low transverse momenta are of particular interest, as the low-x region can be probed. The momentum fraction of the parton x is linked to the transverse momentum p_{T}, the pseudorapidity η and the centre-of-mass energy $\sqrt{s}$ through the formula below [113].

$$x \approx \frac{p_{\text{T}}}{\sqrt{s}} \cdot e^{\pm\eta} \tag{3.20}$$

Therefore, forward/backward jets with low p_{T} can be used to probe the low-x regime. In the discussion on the evolution equations in Sec. 2.2, it was mentioned that the distribution of sea quarks and gluons increases as the proton becomes denser at low-x, effectively increasing the probability for DPS events. Additionally, it was shown in [108] that the DPS pocket formula only upholds if inclusive scenarios in the final state are considered. Therefore, it is opted to examine the inclusive four jet channel in this work.

Inclusive four jet events can be produced in either SPS and DPS, both schematically depicted in Fig. 3.5 on the left and right, respectively. In both the SPS and the DPS case, the four jets with the highest transverse momentum or leading jets will be considered. For a SPS event, all four leading jets are produced in the same interaction, while in the ideal DPS topology two jet pairs, often referred to as a dijet, are created. The differences between the two jet topologies will generally result in stronger correlations between the jets for events produced through SPS compared to DPS. By exploiting the different correlations between the different jets, one can differentiate between the SPS background and the DPS signal.

3.6.4 Choice of Observables

The proper choice of variables is essential in order to study DPS. One aims to measure observables that are expected to display a sensitivity to the topology of DPS events in certain areas of the phase space. The correlations between a dijet or four jets produced in a hard scattering of a DPS event is expected to show distinctly different correlations compared to the dijet or four jets originating from a SPS event. Observable aimed at studying DPS, should therefore exploit these differences as much as possible. Many of such variables have been defined and analyzed in data driven analysis [6–10,12–20,23–27] before. Phenomenological studies have also been performed in the collinear factorization formalism [114–116], as well as phenomenological studies in the k_{T}-factorization formalism [117–119]. Apart from the p_{T} and η spectra of the four leading jets, six observables have been chosen in order to study DPS in four jet production. All of them are listed below.

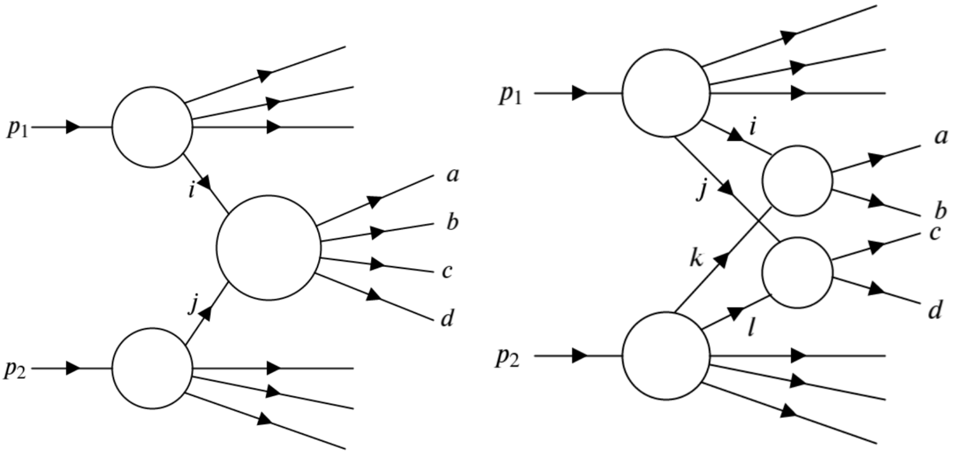


Figure 3.5: A schematic depiction of inclusive four jet production through SPS (left) and DPS (right). In the case of SPS, one hard scattering produces the jets a through d, while two independent hard scatterings create two jets each in the case of DPS. As the two jet pairs are created independently in a DPS event, they are expected to show different correlations compared to four jets originating from a SPS event.

- The azimuthal angular difference between the two jets with the lowest p_{T} or the two softest jets,

  $$\Delta\phi_{\mathrm{soft}} = |\phi_{\mathrm{3rd}} - \phi_{\mathrm{4th}}| \tag{3.21}$$

  The two softest jets are more likely to be in opposite directions in the azimuthal plane due to the conservation of momentum whenever the jets are produced in a DPS event. The difference between their azimuthal angles will approximately be π. Such configurations are referred to as back-to-back configurations. The DPS contribution is therefore expected to peak around $\Delta\phi_{\mathrm{soft}} \to \pi$.

- The minimal combined azimuthal angle of three jets, defined as

  $$\Delta\phi_{\mathrm{3j}}^{\mathrm{min}} = \min\{|\phi_i - \phi_j| + |\phi_j - \phi_k| | i, j, k \in [1, 2, 3, 4], i \neq j \neq k \neq i\} \tag{3.22}$$

  DPS events are more prone to large values of $\Delta\phi_{\mathrm{3j}}^{\mathrm{min}}$ [119], as the two jet pairs are more likely to be in back-to-back configurations in the azimuthal plane. The topology where one jet recoils against the three other jets in the azimuthal plane will occur on average more for SPS events compared to DPS topologies.

- The maximum pseudorapidity difference between two jets,

  $$\Delta\mathrm{Y} = \max\{|\eta_i - \eta_j| | i, j \in [1, 2, 3, 4], i \neq j\} \tag{3.23}$$

  As the maximum separation in pseudorapidity becomes larger, there is a rising possibility that the two jets come from two different interactions or a DPS event.

- The azimuthal angular difference between the jets with the largest pseudorapidity separation,

$$\phi_{\mathrm{ij}} = |\phi_i - \phi_j| \quad \text{for} \quad \Delta\mathrm{Y} = |\eta_i - \eta_j| \tag{3.24}$$

  As the most remote jets, in terms of the pseudorapidity, are more prone to be produced in DPS events, a decorrelation in the distribution of the angular difference of the azimuthal angle of the most remote jets is expected. Whereas the jets are predicted to show stronger correlations, whenever they come from a SPS event.

- The transverse momentum balance between the two softest jets,

$$\Delta p_{\mathrm{T,soft}} = \frac{|\vec{p}_{\mathrm{T,3rd}} + \vec{p}_{\mathrm{T,4th}}|}{|\vec{p}_{\mathrm{T,3rd}}| + |\vec{p}_{\mathrm{T,4th}}|} \tag{3.25}$$

  When the two softest jets originate from DPS, then they are more likely in a back-to-back configuration rendering the value for $\Delta p_{\mathrm{T,soft}}$ small. A more decorrelated picture for SPS is anticipated.

- The azimuthal angle between the hard and soft jet pair,

$$\Delta\mathrm{S} = \arccos\left(\frac{\vec{p}_{\mathrm{T}}(\mathrm{jet}_1, \mathrm{jet}_2) \cdot \vec{p}_{\mathrm{T}}(\mathrm{jet}_3, \mathrm{jet}_4)}{|\vec{p}_{\mathrm{T}}(\mathrm{jet}_1, \mathrm{jet}_2)| \cdot |\vec{p}_{\mathrm{T}}(\mathrm{jet}_3, \mathrm{jet}_4)|}\right) \tag{3.26}$$

  Where $\vec{p}_{\mathrm{T}}(\mathrm{jet_i}, \mathrm{jet_j})$ is the combined transverse momentum vector of the i'th and j'th jet. A more decorrelated picture is expected for DPS compared to SPS, as both jet pairs are independently produced, while the SPS distribution peaks around π, therefore in the tail of $\Delta\mathrm{S} \to 0$ a sensitivity to DPS is projected to be present.

3.7 Monte Carlo Generators

Central to any model trying to describe data or trying to make predictions of certain processes is the MC technique, which refers to the use of random numbers and the use of probability statistics to solve problems. Any model that uses such methods to make physical predictions, is referred to as a MC generator. Many different MC generators exist, which produce different results, depending on the implemented evolution of the partons, the factorization scheme that is used, ... However, the core techniques or operations of these MC generators are the same, i.e. they need to be able to generate numbers according to a certain distribution and they need to be able to calculate integrals. In what follows we will make the distinction between three different groups of models, as we will use the same three groups when discussing the results. The two MPI models implemented by the PYTHIA 8 and HERWIG MC event generators will be discussed at the end of this section.

3.7.1 PYTHIA 8 and HERWIG Models

The PYTHIA 8 and HERWIG MC event generators can produce predictions for a wide variety of processes. Multiple different tunes, i.e. predefined sets of parameters of a generator, have been used with the MC event generators in order to examine the effect of these tunes. All are listed below, along with additional information on the MC event generators themselves.

PYTHIA 8

The PYTHIA 8 MC event generator [92, 93] is a LO MC generator, its 2 → 2 ME is at Born level. The ME is matched with a DGLAP evolution at Next-to-Leading Log (NLL) for the simulation of the parton shower. The implementation of the DGLAP evolution equations yields a strong ordering in p_{T}, i.e. the p_{T} carried by successively emitted partons decreases when moving away from the hard interaction. For the description of the UE, different tunes or sets of predetermined parameters exist. They are obtained by tuning certain parameters in PYTHIA 8 to fit data.

- Three different tunes have been used in PYTHIA 8.240, they are listed below.
    - ⋆ The `CUETP8M1` tune [120, 121] has been obtained by tuning the PYTHIA 8 parameters to CDF data at $\sqrt{s} = 900$ GeV and 1.96 TeV together with CMS data at $\sqrt{s} = 7$ TeV, using the `NNPDF2.3 LO PDF` [122, 123].
    - ⋆ The `CP5` tune [124] is a more recent tune, which uses the newer `NNPDF3.1 nnlo PDF` set [125]. In order to determine the new tune, data at a centre-of-mass energy of $\sqrt{s} = 13$ TeV has been used in addition to data at lower energies.
    - ⋆ The `CDPSTP8S1-4j` tune [121] has been obtained by fitting predictions to two DPS-sensitive observables $\Delta\mathrm{S}$ and $\Delta p_{\mathrm{T,soft}}$ at a centre-of-mass energy of $\sqrt{s} = 7$ TeV, starting from the standard `Tune 4C` [126]. The `CTEQ6L1 PDF` [127] has been used in the tune.

  All three samples have been generated with a minimal value of 15 GeV for the $\hat{p}_{\mathrm{T}}$ parameter, which determines the range of the transverse momenta of each of the produced partons in the rest frame. The last step in the generation of an event is the hadronization where the colored partons coming from the parton shower are converted into colorless hadrons. The PYTHIA 8 MC event generator has implemented the Lund model in this regard.

- A fourth sample has been generated with PYTHIA 8.301 with VINCIA 2.3.01 showering, which has implemented a dipole-antenna showering [128], based on the CDM. The standard PYTHIA 8.301 tune has been used for this sample along with the `NNPDF2.3 LO PDF` set, as no CMS tune for the version 8.301 is available as of yet. The $\hat{p}_{\mathrm{T}}$ parameter is larger than 15 GeV for this sample.

HERWIG

The HERWIG++ and HERWIG 7 event generators [94,95] are both LO MC generators that have implemented a 2 → 2 ME at Born level. For the evolution of partons, the DGLAP evolution equations have been implemented, however, they differ from the PYTHIA 8 MC generator in the ordering of the partons in the parton shower. Whereas PYTHIA 8 uses an ordering in the p_{T} of the emitted partons, HERWIG++ and HERWIG 7 have implemented an angular-ordering. The opening angle of subsequent emitted partons decreases when moving away from the hard interaction. In analogy to PYTHIA 8, different tunes are determined by fitting the MC predictions to data. The different versions of HERWIG in combination with the used tunes are listed below.

- HERWIG++ 2.7.1 has been used in combination with the `CUETHS1` tune [121]. The tune has been obtained in the same manner as the PYTHIA 8 `CUETP8M1` tune, i.e. by tuning parameters in order to make the HERWIG++ predictions fit the CDF data at $\sqrt{s} = 900$ GeV and 1.96 TeV together with CMS data at $\sqrt{s} = 7$ TeV. The `CTEQ6L1 PDF` has been used in the generation of the sample.

- Two tunes have been used in the HERWIG 7.1.5 MC event generator.
  - ⋆ The `SoftTune` is the default tune provided by the HERWIG 7 authors and uses the `MMHT2014lo68cl PDF` [129].
  - ⋆ The `CH3` tune is derived similar to the `CP5` tune: the newer `NNPDF3.1 nnlo PDF` set has been used in combination with data at centre-of-mass energies $\sqrt{s} = 0.9$, 1.96 and 13 TeV.[2]

All HERWIG samples have been generated with a minimal value of 15 GeV for the parameter $\hat{p}_{\mathrm{T}}$. The hadronization is handled by the *cluster model*, as opposed to the *Lund model* in PYTHIA 8.

3.7.2 Multijet Models

A second group of models uses higher order matrix elements. These MC event generators are interfaced to PYTHIA 8.240, PYTHIA 8.301 (with VINCIA), HERWIG 7 or CASCADE 3.0.01-beta1 [130] to include a description of the underlying event, showering and hadronization. Details of the generated event samples are given below.

MadGraph 5

The MADGRAPH 5 (version 2.6.5) [83] is a MC event generator with the ability to compute tree-level and NLO MEs for arbitrary processes. The merging of several LO samples with a parton shower can also be performed by the MADGRAPH 5 event generator, as it has a MLM matching scheme [83, 84] implemented. The three different samples that have been generated with MADGRAPH 5 all use the standard factorization and renormalization scale, they are handled internally. One NLO sample and two merged LO samples have been generated, they are listed below.

- A NLO sample with a $2 \to 2$ ME is interfaced with PYTHIA 8.240, which handles the showering and the hadronization. The `CP5` tune has been used along with the generation cuts of 25 GeV and 20 GeV for the leading and subleading parton, respectively.

- A first LO sample that combines a $2 \to 2$, a $2 \to 3$ and a $2 \to 4$ ME, from now on referred to as $2 \to 2, 3, 4$, has been interfaced with PYTHIA 8.240, using the `CP5` tune. The generation cut on the partons was $H_{\mathrm{T}} > 50$ GeV, where H_{T} is defined as the scalar sum of the transverse momenta of the produced partons.

- A second LO sample with $2 \to 2, 3, 4$ ME has been generated in exactly the same way. The only difference being that parton shower is handled by PYTHIA 8.301 with VINCIA, which uses the standard `NNPDF2.3 LO PDF`.

---

[2] No official documentation exists as of yet, however, the paper is in the final reading stage of the review procedure. The paper has internal handle of CMS-PAS-GEN-19-001.

The two first samples have been generated with the `NNPDF3.1 nnlo PDF`. A more comprehensive overview on the MLM matching procedure of the matrix elements can be found in Appendix A.

PowhegBox

The PowhegBox (version 3633, 2019.02.25) [80–82] is a computer framework for implementing NLO corrections in MC event generators following the powheg method. Each event is built by producing a so-called Born configuration, on which the real phase space is constructed afterward. Two different powheg samples, listed below, have been generated. Both have been interfaced with pythia 8.240 for the showering and hadronization.

- A NLO sample with a 2 → 2 ME [131] has been generated. The factorization and renormalization scale are taken as the minimum p_{T} of the underlying born configuration. In the generation of the sample the minimum p_{T} is set to 10 GeV, which also serves as a generation cut.
- A NLO sample with a 2 → 3 ME [132] has been generated. The generation cut or minimal p_{T} of the underlying Born configuration is set to 10 GeV, while the factorization and renormalization scales are set to $H_{\mathrm{T}}/2$.

The `NNPDF3.1 nnlo PDF` have been used in the generation of both samples.

KaTie

KaTie version 23April2019 [133] is a fairly new LO parton level event generator, which is based on k_{T}-factorization as opposed to collinear factorization. Therefore, it is possible to make on-shell and off-shell calculations, i.e. calculations where the initial state partons have a zero and a non-zero transverse momentum, respectively. KaTie allows for the production of final states with two or more partons as well. In this analysis, it was opted to work with a 2 → 4 ME. All parton level kinematics, used for the generation of samples, are listed below.

- Four partons are generated in the final state with a p_{T} greater than 35 GeV, 30 GeV, 25 GeV, 20 GeV for the leading, subleading, third and fourth leading parton, respectively. Since the cuts are introduced at parton level, the effective thresholds on the resulting hadron level jets are typically 5 to 10 GeV lower.
- The pseudorapidity of all partons is limited to $|\eta_i| < 5.0$.
- The factorization and renormalization scales are set to $H_{\mathrm{T}}/2$.

Four different samples have been generated with these parton level kinematics, i.e. two on-shell and two-off shell samples.

- The first on-shell KaTie sample has been interfaced with pythia 8.240, using the `CP5` tune for the UE description.
- An on-shell KaTie sample has been interfaced with herwig 7.1.5, together with the `CH3` tune.

- Two off-shell KaTie samples have been generated with two different sets of TMD, i.e. the MRW-CT10nlo TMD (MRWTMD) [134] and the PB-NLO-HERAI+II-2018-set2 TMD (PBTMD) [135].

The `NNPDF3.1 nnlo PDF` has been used in the production of both on-shell samples. In the case of the off-shell samples KaTie has been interfaced with Cascade 3.0.01-beta1 [130]. Where Cascade is a MC event generator based on the CCFM evolution equation. It is able to handle particles with a non-zero transverse momentum in the initial state.

3.7.3 SPS+DPS Samples

The PYTHIA 8.240 and KaTie MC event generators both allow the generation of two 2 → 2 matrix elements per event, resulting in a pure DPS sample. In KaTie, the σ_{eff} is a parameter that directly determines the size of the DPS contribution relative to the SPS cross section. A value of 21.3 mb is adopted as in [121]. For pythia 8, it is not possible to set the σ_{eff} as it is determined by the underlying event parameters, and therefore the second 2 → 2 process is simply added with the same kinematic cuts, without any additional scaling of the cross section. Four SPS+DPS samples are used in the analysis in this thesis.

- A pythia 8.240 sample is generated with the `CP5` tune and the `NNPDF3.1 nnlo` PDF.
- The second pythia 8 sample uses the `CDPSP8S1-4j` tune. It is the same sample described in Sec. 3.7.1.
- An on-shell KaTie sample is generated with an explicit DPS contribution, with the exact same generation parameters as the on-shell KaTie LO sample from the multijet models. The showering and hadronization are handled by pythia 8.240 with the `CP5` tune and the `NNPDF3.1 nnlo PDF` set.
- Two off-shell KaTie samples with an explicit DPS contribution are generated using the same TMD as before. Since Cascade can not handle two 2 → 2 matrix elements, non-perturbative corrections have been derived from the on-shell SPS and DPS KaTie samples, and are applied to the off-shell DPS KaTie parton level sample. A more detailed description of the corrections can be found in Appendix B.

It is important to highlight the difference between the SPS+DPS pythia 8 sample using the `CP5` and the `CDPSTP8S1-4j` tunes. The former produces two 2 → 2 MEs, which both have a harder scale compared to the softer UE. The UE parameters have been fitted to DPS-sensitive observables in the latter with the goal that the UE would simulate DPS. Therefore the pythia 8 sample with the `CDPSTP8S1-4j` tune should not need an explicit DPS contribution in order to give a proper description of DPS-sensitive variables.

No SPS+DPS samples have been generated with herwig, as the MC event generator is not able to process two 2 → 2 MEs in the same event file.

3.7.4 Models of MPI

In this analysis, the pythia 8 and herwig MC event generators have been used to handle the description of the UE. Both have implemented different models, which will be discussed below. The study of models of MPI is crucial in order to better understand the proton structure and to possibly improve the MPI models since MPI (and DPS) can play an important role in many physics channels.

MPI in PYTHIA 8

In the discussion of the UE in Sec. 3.3, the existence of MPI from an experimental point of view was conjectured. The divergent behavior of the total partonic cross section σ_{hard} was regulated by the introduction of the minimal exchanged transverse momentum $p_{\text{T,min}}$. In all MPI models implemented in PYTHIA 8, the energy dependence of the regularization parameter $p_{\text{T,min}}$ is modeled by the parameters $p^{\text{ref}}_{\text{T,min}}$, E_{ref} and E_{pow} as given below.

$$p_{\text{T,min}} = p^{\text{ref}}_{\text{T,min}} \left(\frac{E}{E_{\text{ref}}} \right)^{E_{\text{pow}}} \tag{3.27}$$

However, at values of $p_{\text{T,min}} \approx 2–5$ GeV, σ_{hard} becomes larger than the total inelastic cross section σ_{inel}. In order to make sense of this observation, the ratio of the two cross sections can be seen as the average number of partonic interactions in a collision $\bar{n}$, as defined in eq. (3.7).

The simultaneous interactions are assumed to be independent of each other, while the dependence on the impact parameter b is neglected at first. The initial states of the colliding hadrons are assumed to be the exact same for now. Analogue to the definition of $\bar{n}$ in eq. (3.7), the probability of a parton-parton interaction taking place at a given $x_{\text{T}} = 2p_{\text{T}}/\sqrt{s}$ given that two hadrons undergo an inelastic collision, can be defined as

$$f(x_{\text{T}}) = \frac{1}{\sigma_{\text{inel}}(s)} \frac{d\hat{\sigma}}{dx_{\text{T}}} \tag{3.28}$$

All other objects are defined as before. In the framework of the evolution equations and parton showers, the Sudakov form factor is defined as the probability of a parton i evolving from an initial scale to a different scale without the occurrence of a branching, see eq. 3.5. In the same way, the probability of the first and hardest interaction to occur at a scale $x_{\text{T},1}$ can be written as

$$P(x_{\text{T},1}) = f(x_{\text{T},1}) \exp \left(\int_{x_{\text{T},1}}^{1} f(x'_{\text{T}}) dx'_{\text{T}} \right) \tag{3.29}$$

Through iteration, one can find the probability of an i'th scattering taking place at the scale $x_{\text{T,i}}$ which is given below

$$P(x_{\text{T,i}}) = \frac{f(x_{\text{T,i}})}{(i-1)!} \left(\int_{x_{\text{T,i}}}^{1} f(x'_{\text{T}}) dx'_{\text{T}} \right)^{i-1} \cdot \exp \left(\int_{x_{\text{T},1}}^{1} f(x'_{\text{T}}) dx'_{\text{T}} \right) \tag{3.30}$$

The scales of all the interactions are ordered from hardest to softest or

$$1 > x_{\mathrm{T,1}} > x_{\mathrm{T,2}} > \ldots > x_{\mathrm{T,i-1}} > x_{\mathrm{T,i}} > x_{\mathrm{T,min}} \tag{3.31}$$

No additional interactions can occur once the scale drops below the threshold $x_{\mathrm{T,min}} = 2p_{\mathrm{T,min}}/\sqrt{s}$, determined by the minimal exchanged transverse momentum. For the hardest scattering, the normal PDF should be used. However, for MPI produced through this mechanism, the PDF should be rescaled in order to take the conservation of four-momentum into account which is performed by rescaling the $x_{\mathrm{T,i}}$ of the i'th scattering as

$$x'_{\mathrm{T,i}} = \frac{1}{1 - \sum_{j=1}^{i-1} x_{\mathrm{T,j}}} \tag{3.32}$$

The rescaled $x'_{\mathrm{T,i}}$ takes the energy used in all previous interactions into account. After each interaction and scaling, the size of the available phase space is reduced until no room and energy are left for any additional interactions.

In the discussion on the hard scattering in Sec. 3.2, ISR and FSR were added in order to obtain a more complete picture of the interaction. The same has to be done for each MPI, as the partons can emit radiation before and after the interaction. In this regard, the formalism described above, is implemented in an interleaved evolution of the hard interaction, MPI, ISR and FSR. The master equation [136] is given below.

$$\begin{aligned}\frac{dP}{dp_{\mathrm{T,i}}} = &\left(\frac{dP_{\mathrm{MPI}}}{dp_{\mathrm{T,i}}} + \sum \frac{dP_{\mathrm{ISR}}}{dp_{\mathrm{T,i}}} + \sum \frac{dP_{\mathrm{FSR}}}{dp_{\mathrm{T,i}}}\right) \\ &\times \exp\left(-\int_{p_{\mathrm{T,i}}}^{p_{T,i-1}} \left(\frac{dP_{\mathrm{MPI}}}{dp'_{\mathrm{T}}} + \sum \frac{dP_{\mathrm{ISR}}}{dp'_{\mathrm{T}}} + \sum \frac{dP_{\mathrm{FSR}}}{dp'_{\mathrm{T}}}\right) dp'_{\mathrm{T}}\right)\end{aligned} \tag{3.33}$$

Where $p_{T,i}$ is the transverse momentum of the i'th interaction or parton branching. The formula takes a similar form compared to Sudakov form factor and expresses the probability that no interaction or parton branching occurs between the $(i-1)$'th interaction/parton branching and the current one. The inclusion of FSR in the master equation, allows for the concept of rescattering [137] to be added the description of MPI. An example of a rescattering process is given in Fig. 3.6; an outgoing parton from one interaction serves as the initial or incoming parton of another scattering.

Up until this point, the incoming hadrons are assumed to be identical for every collision, however, in reality this is not the case. The differences in the centrality of the collision are modeled as a function of the impact parameter b. Whenever the parameter b is small (large), a large (small) overlap is present between the colliding hadrons, effectively increasing (decreasing) the probability for MPI. As the matter distribution for all partons inside the hadron is assumed to display a spherical symmetry, the matter distribution used to quantify the overlap of the colliding hadrons is represented by a double Gaussian, given below.

$$\rho(r) \sim \frac{1-\beta}{a_1^3} \exp\left(-\frac{r^2}{a_1^2}\right) + \frac{\beta}{a_2^3} \exp\left(-\frac{r^2}{a_2^2}\right) \tag{3.34}$$

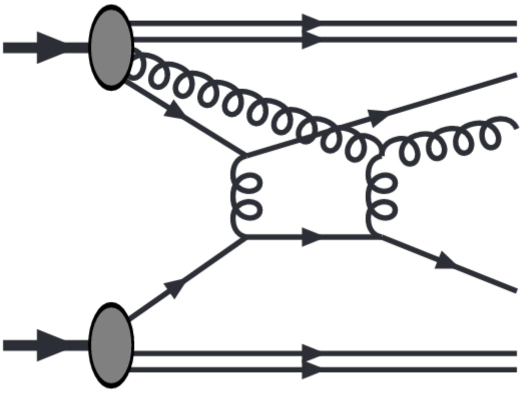


Figure 3.6: A Feynman diagram of a DPS event where one of the outgoing partons of one interaction serves as an incoming parton or initial state of the second interaction, taken from [137]. Such an occurence is referred to as the rescattering of the parton.

The distribution corresponds to a small core region which contains a fraction β of the total hadronic matter within a radius a_2. The smaller core is embedded in a larger hadron with radius a_1.

A last ingredient that is added to the MPI model implemented in PYTHIA 8, is the so-called colour reconnection. As partons carry a colour charge, the colour reconnection allows for colour strings originating from different interactions to connect and exchange colour information. In MPI processes, many colour strings are often created that can overlap in the physical space, whereafter they would naturally reconnect with each other. The reconnection procedure uses the probability of a system reconnecting to a harder one defined as

$$P = \frac{p_{\mathrm{T,Rec}}^2}{p_{\mathrm{T,Rec}}^2 + p_{\mathrm{T}}^2} \quad \text{with} \quad p_{\mathrm{T,Rec}} = R \cdot p_{\mathrm{T,min}} \tag{3.35}$$

Where $p_{\mathrm{T,min}}$ is the parameter regulating the cross section σ_{hard} and R is a parameter of the model.

MPI in HERWIG

The same MPI mechanism has been implemented in both HERWIG++ and HERWIG 7 [138]. The model is based on the same observation as discussed in Sec. 3.3 but for dijet production: the inclusive cross section for dijet production σ_{inc} will outgrow the total cross section at high enough energies. The observation leads to the conjecture of the occurrence of MPI in a hadron collision. Two main approximations are made to obtain the MPI model. On the one hand, the different scatterings are assumed to be independent at a fixed impact parameter $\vec{b}$. On the other hand, it is reckoned that the

PDF factorizes with respect to the $\vec{b}$ and x dependence. With these two assumptions in mind, the average number of interactions in hadron collision $\bar{n}$ can be expressed as

$$\bar{n}(\vec{b}, s) = A(\vec{b}, \mu^2) \cdot \sigma_{\text{inc}}(s, p_{\text{T,min}}) \tag{3.36}$$

Where the function $A(\vec{b}, \mu^2)$ describes the overlap between the two incoming hadrons, the parameter μ is interpreted as the inverse proton radius and $p_{\text{T,min}}$ is the same cutoff used to regulate the cross section as before. Below the threshold a non-perturbative model of MPI is implemented, see [138] for more details. In order to improve the model, colour reconnection has been implemented as was done in the MPI model in PYTHIA 8.

3.8 Generator Level Study of DPS

A first generator level study of the observables defined in Sec. 3.6.4 is performed with the on-shell KATIE model, interfaced with the PYTHIA 8 MC event generator. The resulting SPS and DPS distributions on hadron level are shown in Fig. 3.7, Fig. 3.8 and Fig. 3.9 for the p_{T} spectra, the pseudorapidity spectra and the six DPS-sensitive variables, respectively. The sum of the SPS and DPS samples (SPS+DPS) is also calculated and depicted on the figure. On the bottom panel of the figures, the ratio of the SPS+DPS distribution over the SPS curve is presented, indicating in which regions of the phase space a larger DPS contribution is expected.

All the plots except for the ΔS distribution have been obtained by imposing the cuts of 35-30-25-20 GeV on the four leading jets. The ΔS observable has been produced for cuts of 50-30-30-30 GeV. The reason for the two sets of cuts will be made clear later on in the analysis.

All the p_{T} spectra of the four jets, depicted in Fig. 3.7, show the largest contribution of DPS for low values of p_{T} since the secondary scattering of DPS events is generally softer as the first hard interaction. The pseudorapidity spectra in Fig. 3.8 display a more homogeneous distribution of the DPS contributions, with the exception of the most forward and backward bins. Jets in the most remote areas of the detector are more likely to have a lower p_{T} compared to jets in the central region of the detector. Therefore a larger DPS contribution is expected in the forward and backward region. All other variables shown in Fig. 3.9, are sensitive to DPS, i.e. show an increase in the DPS contribution to the total cross section in the exact regions, as detailed in Sec. 3.6.4, with the exception of the observable ϕ_{ij}. The variable ϕ_{ij} displays a more homogeneous contribution over the whole interval, with a slight rise towards zero and π.

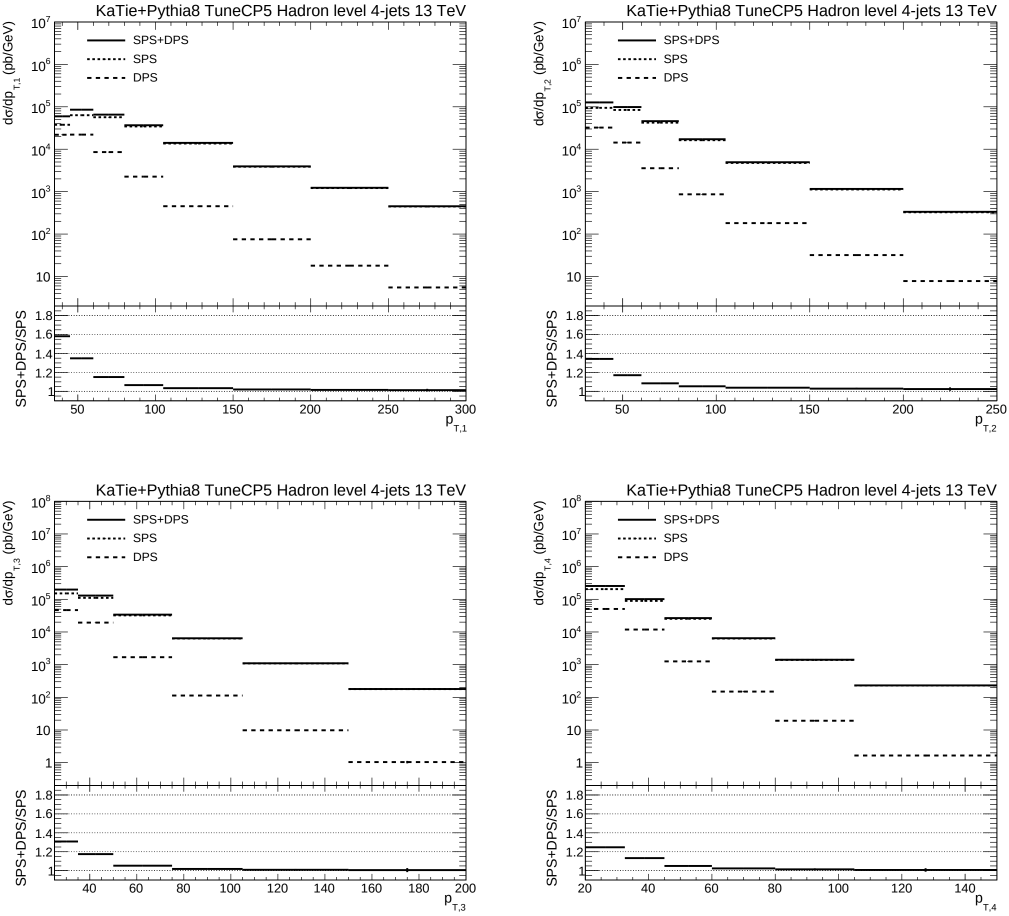


Figure 3.7: The generator level results for the on-shell KaTie model, interfaced with PYTHIA 8, for the p_{T} spectra for the leading (top left), subleading (top right), third leading (bottom left) and fourth leading (bottom right) jets. The SPS, DPS and total distributions are shown, together with the ratio of the total distribution compared to the pure SPS sample.

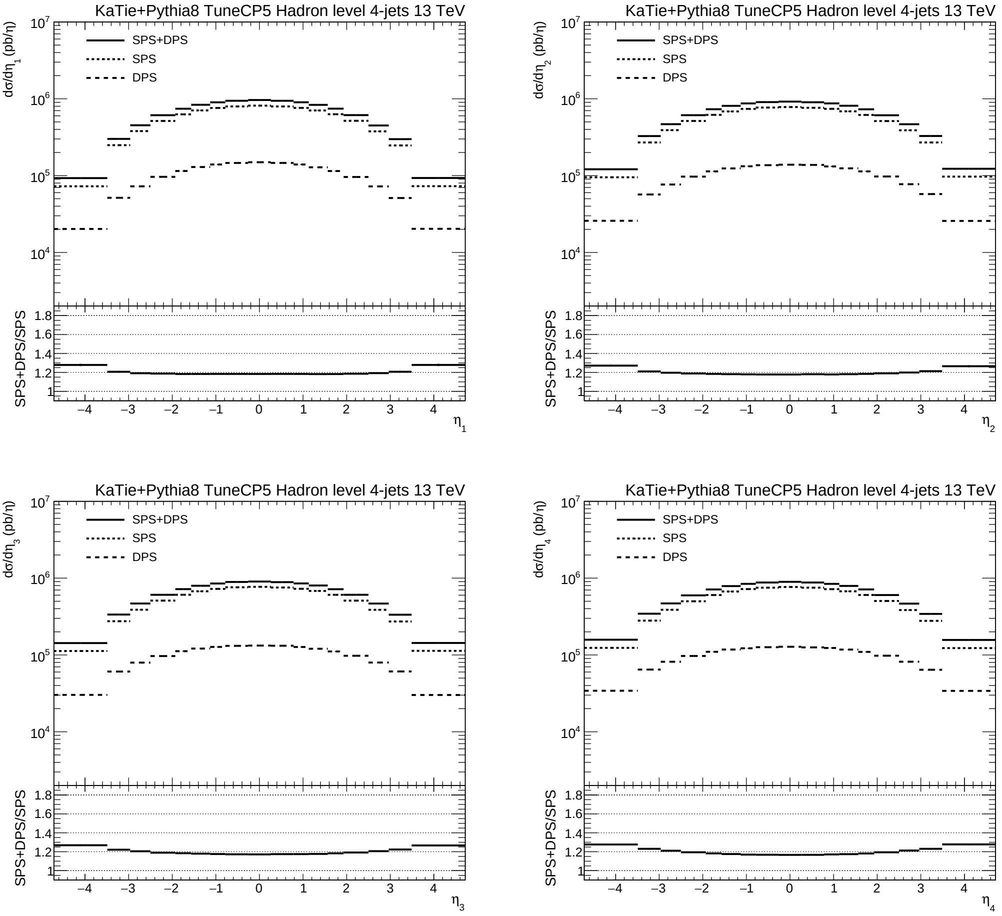


Figure 3.8: The generator level results for the on-shell KaTie model, interfaced with PYTHIA 8, for the η spectra for the leading (top left), subleading (top right), third leading (bottom left) and fourth leading (bottom right) jets. The SPS, DPS and total distributions are shown, together with the ratio of the total distribution compared to the pure SPS sample.

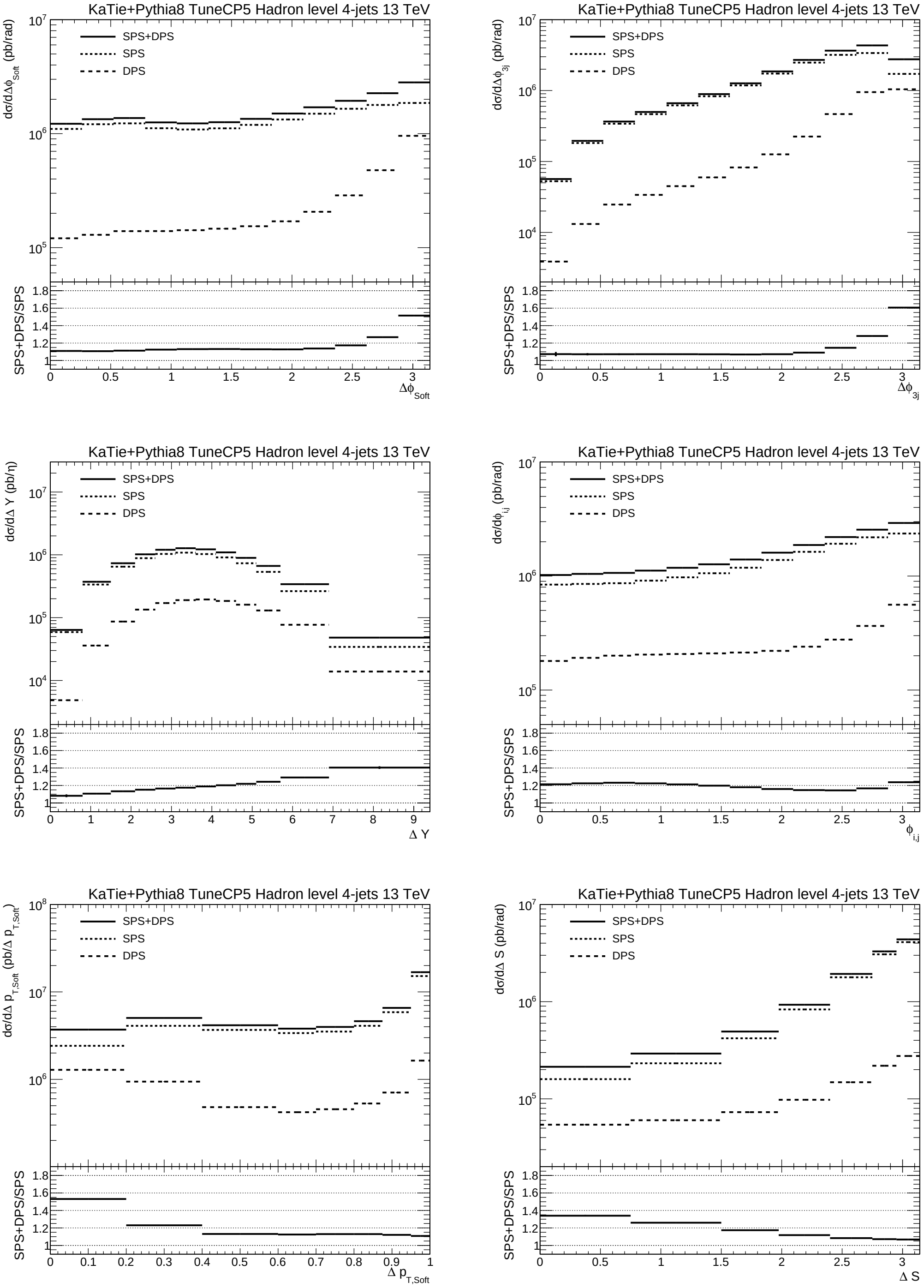


Figure 3.9: The generator level results for the on-shell KaTie model, interfaced with PYTHIA 8. The ΔS variable (bottom right) is subjected to different selection criteria, i.e. the cuts on the p_{T} of the jets are 50-30-30-30 GeV, respectively.

Chapter 4

The Large Hadron Collider and the Compact Muon Solenoid Experiment

In this chapter, the Large Hadron Collider (LHC) and the Compact Muon Solenoid (CMS) experiment are described. The machinery relevant to this work and the manner in which collisions of interest are detected and stored will be highlighted.

4.1 The Large Hadron Collider

The LHC [139] is a ring accelerator with a circumference of 26.7 km at a depth ranging from 45 m to 170 m below the surface, operating at Centre Européen pour la Recherche Nucléaire (CERN), located near Geneva. The LHC consists of two rings in which counter-rotating beams of protons or heavy ions are accelerated. At different interaction points where different experiments are located, the beams are crossed.

Apart from the type of collider, linear or circular, and the colliding particles, the centre-of-mass energy of the collision and the instantaneous luminosity characterize a collider. The former is of the TeV scale, more precisely $\sqrt{s} = 13$ TeV for the 2016 run period. The latter is the parameter that describes the number of particles in the beam at the interaction point per unit of transverse area and per unit of time, and is defined as

$$L = \frac{N_{\mathrm{b}}^2 n_{\mathrm{b}} f_{\mathrm{rev}}}{\pi \epsilon_{\mathrm{n}} \beta^*} F \tag{4.1}$$

Where $f_{\mathrm{rev}} = c/r$ is the revolution frequency, equal to 11.25 kHz for the LHC and ϵ_{n} is the normalized transverse beam emittance. Since the beams cross under a certain angle at the interaction point, the geometric luminosity reduction factor function F is introduced, while the β^* function represents the focus of the beam at the collision point. Both F and β^* describe the interaction point. Lastly, the variables N_{b} and n_{b} characterize the proton beams, i.e. protons are bundled together in bunches when circulating. The variables N_b and n_b represent the number of protons per bunch and the number of bunches, respectively. A more detailed description can be found in [139].

The LHC has been designed to run at a centre-of-mass energy of 14 TeV along with a peak instantaneous luminosity $L \sim 10^{34}$ cm^{-2}s^{-1} and a maximum of 2808 proton bunches, where each bunch consists of approximately $1.18 \cdot 10^{11}$ protons. However, during the year 2016 the centre-of-mass energy was 13 TeV, while the maximum number of proton bunches achieved was 2220. The number of bunch crossings at the interaction points were at a frequency equal to 40 MHz, corresponding to a separation between the bunches of 25 ns or 7.5 meters, while the bunches themselves were approximately 8 cm long and had a radius of 20 μm.

Multiple experiments are located along the LHC at different interaction point. The two largest experiments are the A Toroidal LHC ApparatuS (ATLAS) [140] and CMS [141, 142] experiments, both aim to achieve the peak luminosity of $L \sim 10^{34}$ cm^{-2}s^{-1}. Two low luminosity experiments Large Hadron Collider beauty (LHCb) [143] and ToTal Elastic and diffractive cross section Measurement (TOTEM) [144] aim at a peak luminosity of $L \sim 10^{32}$ cm^{-2}s^{-1} and $L \sim 2 \times 10^{29}$ cm^{-2}s^{-1}, respectively. A fifth experiment dedicated to lead-lead ion collisions has been named A Large Ion Collider Experiment (ALICE) [145] with a peak luminosity of $L \sim 10^{27}$ cm^{-2}s^{-1}. Since the cross sections of a process is measured in barn $= 10^{-28}$ m^2, the instantaneous luminosity is generally expressed in units of barn^{-1}. The instantaneous luminosity and the cross section of a process are related in the following way

$$L = \frac{1}{\sigma} \cdot \frac{dN}{dt} \tag{4.2}$$

Where dN/dt is the rate of the process with the cross section σ.

Before measurements can be performed at the different experiments, the protons need to be accelerated and injected into the LHC, which happens in different stages as the final energy of each proton beam is 6.5 TeV [139]. A schematic overview of all the machines involved in the injection procedure is depicted in Fig. 4.1. The source of the protons is a hydrogen gas. An electric field is used to ionize the hydrogen gas. The LINAC2 is the first machine in the accelerator chain and brings the protons to an energy of 50 MeV. Next, the protons enter the proton synchrotron booster, which is a small four-ring synchrotron with a radius of 25 m. Once the proton synchrotron booster has brought the energy of the proton beams up to 1.4 GeV, they are transferred to the proton synchrotron, a synchrotron with a 100 m radius. In the proton synchrotron the final bunch structure of the protons is obtained, along with a beam energy equal to 25 GeV. In the final before injection into the LHC, the proton beams are passed on to the super proton synchrotron, a synchrotron with a radius of 1 km. Once the beam energy reaches 450 GeV, the proton beams are injected into the LHC, where they are accelerated up to their final beam energy of 6.5 TeV. The LHC uses superconductive magnets, i.e. 1232 dipole and 392 quadrupole magnets, to bend and focus the beams, respectively, whilst operating at a temperature of 1.9 K.

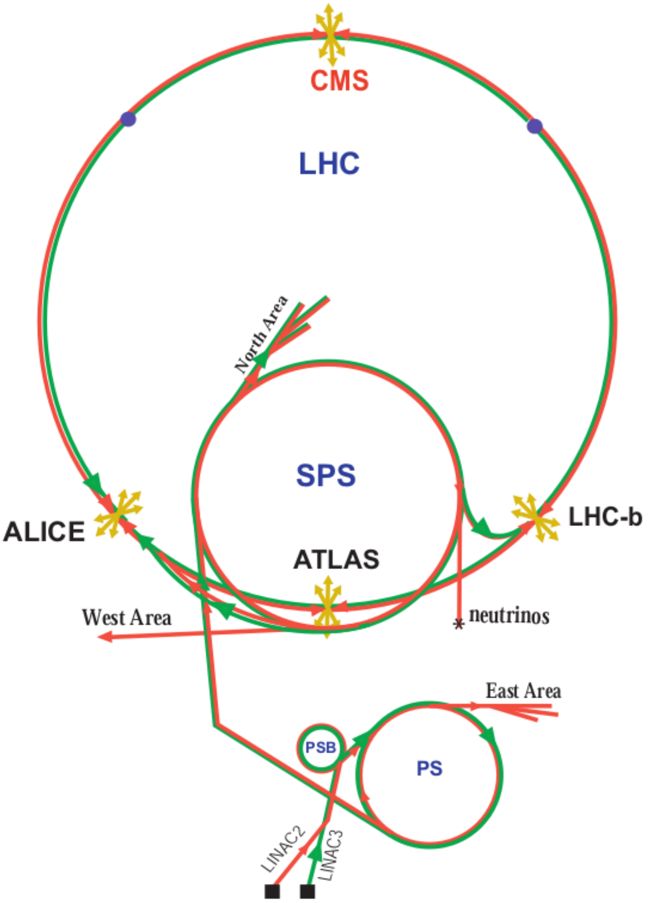


Figure 4.1: An overview of the whole accelerator complex at CERN. The proton beams are accelerated by several facilities in succession before injection into the LHC. The different interaction points and experiments are also shown. Taken from [146].

4.2 The CMS Detector

The CMS detector [141, 142] is a multi purpose detector located at the interaction point 5 (IP5) of the LHC. Its central feature is a 6 m diameter superconducting solenoid, which is able to generate a 3.8 T magnetic field. The CMS detector has a cylindrical symmetry with an overall length of 28.7 m, a diameter of 15 m and a total weight of 14,000 tonnes. The proton interaction point is situated at the centre of the detector. The tracker is the sub-detector closest to the interaction point, around which the electromagnetic and hadronic calorimeters are situated. All three sub-detectors are situated inside the solenoid, except for some sub-systems of the hadronic calorimeter. While four layers of muon chambers combined with the steel return yoke encompass the solenoid in the barrel region and the endcaps. The structure of the detector is schematically depicted in

Fig. 4.2.

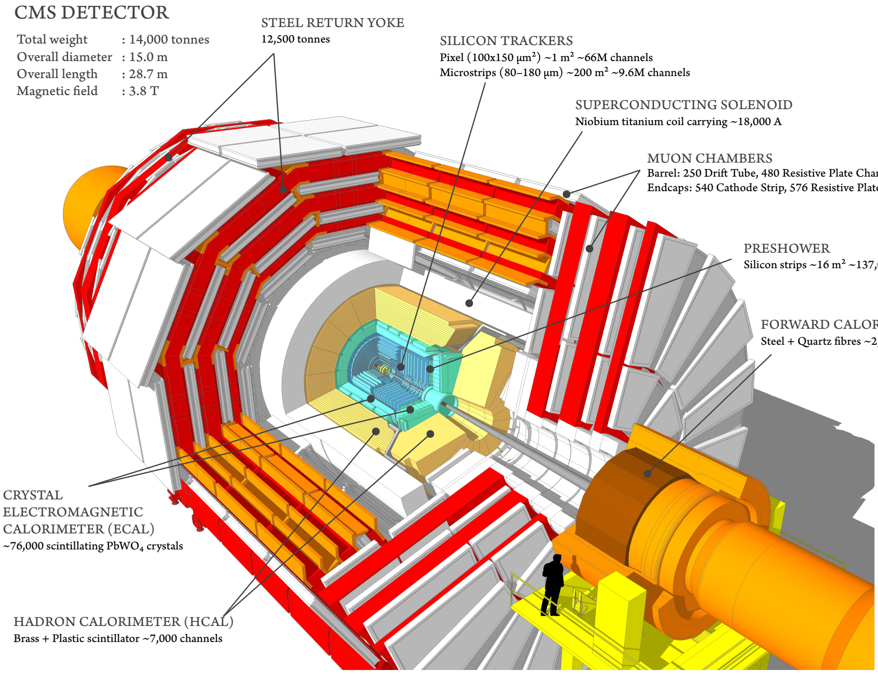


Figure 4.2: An overview of the CMS detector at the LHC. All sub-detectors are proportionally displayed in the sketch. Taken from [147].

4.2.1 Coordinate System of the CMS Detector

The CMS coordinate system is right-handed, where the x-axis points towards the centre of the LHC ring, while the y-axis points upward. Therefore, the z-axis is aimed along the anticlockwise direction of the beam. In the xy-plane, radial coordinates are used where ϕ is defined as the azimuthal angle and ρ as the radial coordinate. Instead of the polar angle θ defined in the ρz-plane, it is often preferred to use the pseudorapidity η, defined in eq. (3.13). Objects are referred to as "central" when $|\eta| < 2.5$, while the adjective "forward" is used for objects with $|\eta| > 2.5$. Generally objects are measured in quantities of η, ϕ and p_{T}.

4.2.2 The CMS Tracker

At the heart of the CMS detector, fully immersed in the magnetic field provided by the CMS solenoid, lies the CMS tracker [148, 149]. The CMS tracker is a cylindrical sub-detector with a length of 5.8 m and a diameter of 2.5 m, while covering a pseudorapidity region up to $|\eta| = 2.5$. A sketch of the tracker is shown in Fig. 4.3. The sub-detector is designed to provide precision measurements of charged particles with a p_{T} as low as

0.1 GeV, while having a high granularity and a fast response time, as ~ 1000 particles are expected to traverse the tracker per bunch crossing. A high resolution is essential in the reconstruction process of interaction vertices, as well as for the identification of heavy quark flavours. Furthermore, the information provided by the tracker is crucial for a proper jet reconstruction and allows for the identification of electrons and muons when combined with the calorimeters and muon chambers, respectively. Apart from the offline event reconstruction, the tracker is also used in the CMS High-Level Trigger (HLT), discussed in Sec. 4.2.6.

The pixel detector, located at the centre of the CMS tracker, consists of 1,440 modules spanning a total area of 2 m^2 with a total of 66 million pixels. Three cylindrical barrel layers at radii of 4.4, 7.3 and 10.2 cm, and two endcap disks at 34.5 and 46.5 cm of each side of the interaction point make up the pixel detector. The resolution of the pixel detector is 10 and 20-40 μm in the transverse and longitudinal direction, respectively, while a third coordinate is provided by the sensor plane position.

The silicone strip tracker encapsulates the pixel detector and is divided into four subsystems, consisting 15,148 silicon modules, which have in total 9.3 million strips with a total active area of 198 m^2. The Tracker Inner Barrel (TIB) and the Tracker Outer Barrel (TOB) form the central part of the strip tracker, where the strips are aligned cylindrical along the beam pipe or x-axis. The TIB is composed of four barrel layers and provides a resolution of approximately 13-38 μm on the position in the azimuthal plane, while the TOB is made up from 6 layers with a slightly lower resolution of 18-47 μm in the azimuthal plane. The endcap region is covered by two subsystems as well. The Tracker Inner Disks (TID) and the Tracker EndCaps (TEC) consist of three and nine disks, respectively, where the disks themselves contain three to seven concentric rings of strip modules. The resolution in the azimuthal plane of the TID and TEC are of the same range as the resolution of the TIB and TOB, respectively.

The performance of the tracker as a whole has been studied and the results have been described in [149]. The resolution has been analyzed for simulated muons and charged pions, using the five track parameters d_0, z_0, ϕ, $\cot\theta$ and p_{T}. The parameters are defined at the impact point or the point of closest approach to the beam axis, where $d_0 = x_0 \sin\phi - y_0 \cos\phi$ is the distance from the beam axis to the impact point in the azimuthal plane. In Fig. 4.4 the resolution of all five parameters is shown as a function of η for single isolated pions with a p_{T} equal to 1, 10 and 100 GeV. The resolution in general is calculated by examining the residual differences between the detector and generator level observables. The full symbols correspond to the half-width of the distribution containing 68%, while the open symbols represent the same half-width where 90% has been used instead. The latter shows the impact of extreme values more clearly. The resolution of the transverse impact parameter improves from 90 μm at 1 GeV to 25 μm at 10 GeV, while for the longitudinal impact parameter an improvement of 150 to 45 μm is observed from 1 to 10 GeV, respectively, both for central tracks. The resolution in p_{T} never exceeds 4% in the central region for tracks with a p_{T} ranging up to 100 GeV. Lastly, the resolution of ϕ and $\cot\eta$ are well behaved and remain smaller than 0.01 in the central region of the tracker.

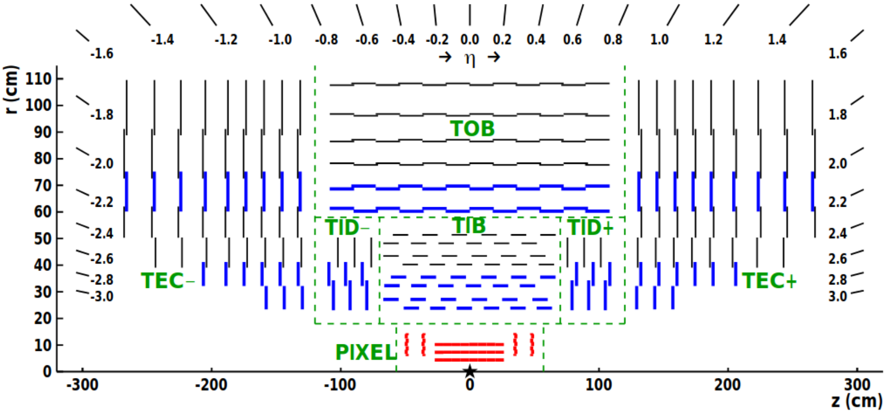


Figure 4.3: An overview of the CMS tracker in the ρz-plane. The collision point is indicated by a star. The pixel detector is represented by the red lines, while the black and blue lines indicate single and double sided strip modules. The green dashed lines indicate the boundaries of the different subsystems. Taken from [149].

4.2.3 The CMS Magnet

The central feature of the CMS detector is the CMS magnet [150], it has a length of 13 m and a diameter of 6 m. The 220 tons superconducting solenoid, operating at a temperature of 4.5 K, is able to generate a uniform magnetic field of up to 4 T by running a current of 18,000 A through the niobium titanium coil. A 12,500 tons weighing steel return yoke consisting of five barrel rings and two endcaps, each made up of 3 layers, close the magnetic field. The field is used to bend charged particles since the curvature is related to the transverse momentum of said particles. Due to the sheer force of the magnetic field in combination with the high-precision position measurements of the tracker, even the momentum of the high-energy particles created in collisions can be accurately measured. A second advantage of the presence of a strong magnetic field is that it allows for a relative "compact" design of the detector.

4.2.4 The CMS Calorimeter System

Around the tracker and for the most part inside the CMS solenoid, the CMS calorimeter system is situated. Contrary to the tracker system, the calorimeter system relies on the destructive measurement of particles to determine their respective energies, which is achieved by initiating particle showers in the detector material. The distinction between electromagnetic and hadronic calorimetry is made since showers produced by an electron, a positron or a photon display different characteristics compared to hadronic showers.

Pair production and bremsstrahlung are the two dominant processes in the creation of the electromagnetic showers. The former describes the splitting of photons in electron-positron pairs through the interaction with a nucleus or electron in order to conserve momentum. The latter process occurs when a high-energy electron or positron is decelerated through the emission of photons, which in turn can split into electron-positron

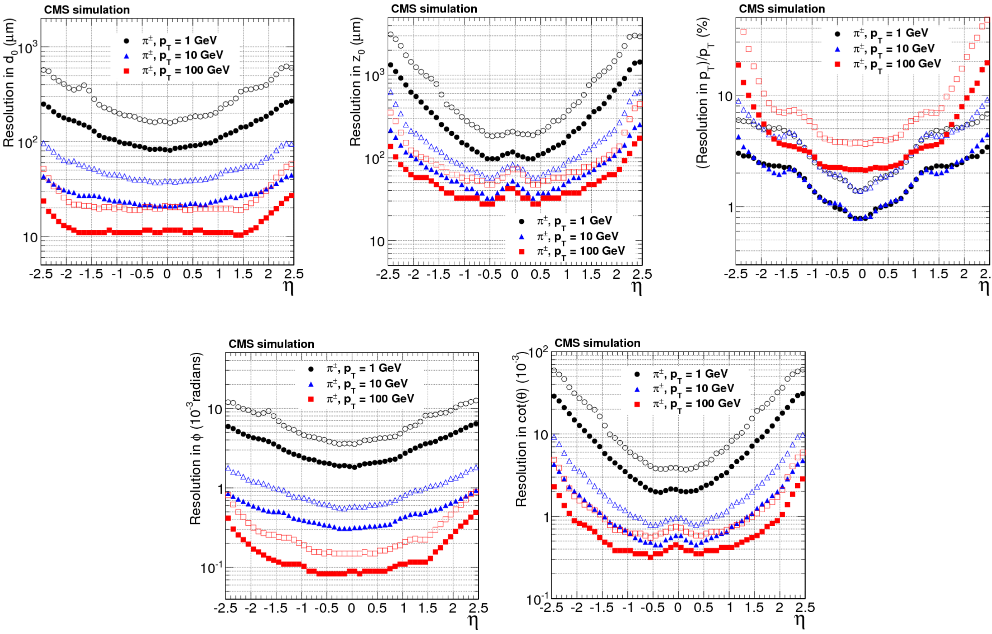


Figure 4.4: Resolution, as a function of η, in the five track parameters for charged pions with a p_T equal to 1, 10 and 100 GeV. From top to bottom and left to right: the transverse impact parameters d_0, the longitudinal impact parameters z_0, the transverse momentum p_T, the azimuthal angle ϕ and the cotangent of the polar angle $\cot\theta$. The solid symbols correspond to the half-width that encompasses 68% of the residuals, while 90% of the residuals is used for the open symbols. Taken from [149]

pairs. The interplay of the two processes continues until the energy of the photons does not meet the pair production threshold anymore. Other energy loss processes of the electrons and positrons will now start to dominate instead of bremsstrahlung. The electromagnetic showers are characterized by two material specific constants. On the one hand, the radiation length X_0 is the average distance needed for an electron to reduce its initial energy by a factor of $1/e$. On the other hand, the Molière radius R_M is a measure for the transverse profile of the electromagnetic shower, i.e. the radius defines a cylinder containing on average 90% of the total energy deposition of the shower.

The hadronic showers are formed, contrary to electromagnetic showers, through inelastic scatterings of hadrons in the detector material. The incoming particles lose their energy through nuclear excitation and hadron production. One can assume that all final particles are pions, one third of which are neutral, while the remainder are charged pions. The former decays in two photons, forming the electromagnetic part of the hadronic shower. The hadronic equivalent of the radiation length X_0 is the interaction length λ. In other words: the distance needed for an incoming hadron to lose a factor of $1/e$ of its energy is equal to λ. On average hadronic showers are less compact and longer compared to electromagnetic showers, which is reflected in the fact that λ is generally larger than

X_0 for a given material.

The CMS Electromagnetic Calorimeter

The CMS Electromagnetic CALorimeter (ECAL) [151,152] is a homogeneous calorimeter, comprised of the ECAL Barrel (EB) and ECAL Endcaps (EE), each containing 61,200 and 7324 lead tungstate ($PbWO_4$) scintillating crystals, respectively. In order to boost the photon identification capabilities of the endcap crystals, a ECAL preShower dector (ES) has been placed in front of them. The $PbWO_4$ crystals have a high density (8.28 g/cm^3), a short radiation length (0.83 cm) and a small Molière radius (2.2 cm). The construction of a "compact" calorimeter with a high granularity was possible due to the fast response time and a good radiation tolerance of the $PbWO_4$ crystals. The scintillation light, originating from the electrons and positrons in the electromagnetic shower, is transmitted through total internal reflection and is caught by photodetectors.

The EB crystals are 23 cm long with a 2.2 cm × 2.2 cm face, while the EE crystals have a 2.85 cm × 2.85 cm face and a 22 cm length. The surface of the crystals in both the EE and EB correspond to a 0.0174 × 0.0174 coverage in $\Delta\eta \times \Delta\phi$. The barrel region has an inner radius of 129 cm and is comprised of 18 supermodules on each side of the interaction point. The supermodules cover the range of $|\eta| < 1.479$ and consist of four modules along the η direction. The EE covers the range of $|\eta| \in [1.653, 3.0]$, while being at a distance of 314 cm from the interaction point. The crystals are grouped in the EE in units of 5 × 5 to form an approximately circular shape. The ES has been placed in front of the EE and cover the range of $|\eta| \in [1.653, 2.6]$. The silicon strip sensors of the EE allow for a more exact position measurement and an enhanced photon identification. A schematic overview of the ECAL is given in Fig. 4.5.

The energy resolution for electrons has been measured and is parametrized by the function below [152].

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E}} + \frac{12\%}{E} + 0.3\% \tag{4.3}$$

Where the energy E is measured in GeV. The first term is the stochastic term, receiving contributions from the shower containment, the number of photo-electrons and fluctuations in the gain process. The second term represents the contribution of the noise due to the electronics. While the last and constant term is due to energy leakage, non-uniformity of the light collection and calibration errors. For high energy electromagnetic showers, the last term will dominate the energy resolution.

The CMS Hadronic Calorimeter

The CMS Hadronic CALorimeter (HCAL) [153] is a sampling calorimeter and is comprised of four subsystems: the HCAL Barrel (HB), the HCAL Endcaps (HE), the HCAL Outer detector (HO) and the HCAL Forward detector (HF). Unlike the ECAL, not all subsystems lie within the superconducting solenoid, i.e. only the HB and the HE are inside the solenoid. A schematic overview of the HCAL detector is shown in Fig. 4.6.

The HB covers the pseudorapidity range up to $|\eta| = 1.4$. Along the beam line, the HB has a cylindrical symmetry and is built of 18 wedges, each covering an azimuthal angle of 20 degrees. In turn, all wedges are divided into four 5 degree sectors in ϕ and are made

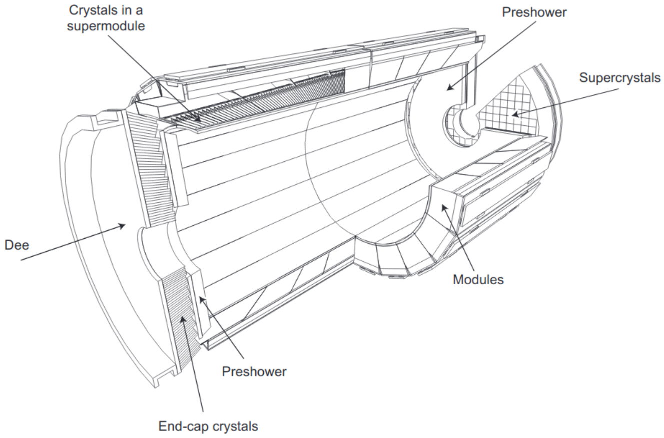


Figure 4.5: An overview of the CMS ECAL. The structure of the modules and supermodules of the barrel region is shown, together with the endcaps and the preshower detectors. Taken from [152].

up from 14 flat brass absorber layers in between two steel plates. A second segmentation in η is made, resulting in an overall division of the detector in $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. The hadronic interaction length λ ranges from 5.8 at the centre, up to a value of 10λ at $\eta = 1.2$.

The HE is made of brass disks, bedded between scintillator wedges and has the same azimuthal segmentation of 20 degrees in ϕ. The pseudorapidity range of the HE is $|\eta| \in [1.3, 3.0]$. Both the HB and the HE cover the pseudorapidity range $|\eta| \in [1.3, 1.4]$ to minimize the effect of an uninstrumented gap at a polar angle of 57 degrees. The HE has the same granularity as the HB except for $|\eta| \geq 1.74$, where the ϕ segmentation is 10 degrees or 0.175 rad, while the segmentation in η varies from 0.09 up to 0.35.

The HO sub-detector lies outside the solenoid and consists of five rings, each covering a different rapidity range. The central ring covers $|\eta| \leq 0.35$, the rapidity range between 0.35 and 0.87 is contained in the rings ±1 , while the rings ±2 extend the rapidity range from 0.87 up to 1.26. The HO is designed to catch any possible energy leakage from the HB from penetrating hadron showers, while having a 12-fold symmetry, where each component covers a 30 degree angle in ϕ. As each component is divided into six 5 degree sectors and as the HO has the same rapidity segmentation, the overall segmentation of the HO in $\Delta\eta \times \Delta\phi$ is the same as for the HB or 0.087×0.087

The last subsystem is the HF, covering a range of $|\eta|$ between 3.0 and 5.2, and is located further outward along the beamline at a distance of 11.15 m from the interaction point. The HF is constructed from 20 degree wedges, which are in turn divided in two 10 degree sectors in ϕ. The HF is constructed from quartz fiber interleaved with steel absorber layers due to the high flux of particles in the forward directions. Both electromagnetic and hadronic showers can be detected by the HF, where Cherenkov radiation forms the basis of signal generation. The segmentation is 0.175×0.175 except for $|\eta| > 4.7$, where it is 0.175×0.35.

The performance in terms of the synchronization and timing is detailed in [154], while the energy resolution of the HCAL subsystems has been studied in [155,156]. The energy resolution is parametrized by the sum of a stochastic and constant term, given in the function below.

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} + b \tag{4.4}$$

The values of a and b for the HB and HE are similar and equal to 8.47% and 7.4%, respectively. The parametrization for HF is more complex and differentiates between sections and the type of shower. For a more detailed discussion on the parametrization of the energy response in HF and the values of a and b, we refer to [156].

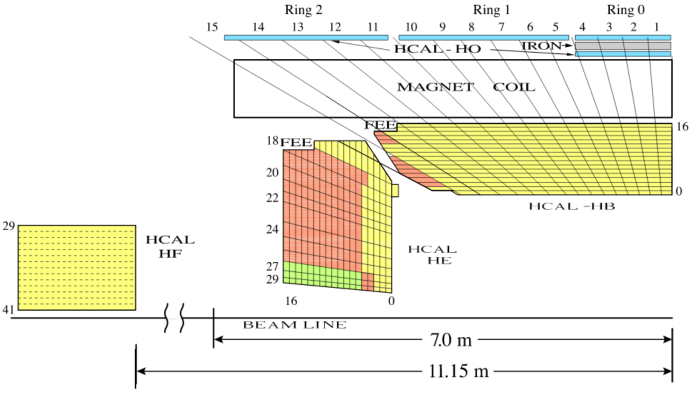


Figure 4.6: An overview (quarter slice) of the CMS HCAL. The position of all sub-detectors is shown, where FEE stands for front end electronics, indicated for HB and HE. The signals of the tower segments with the same color are added optically to provide a longitudinal segmentation. Taken from [155].

4.2.5 The Muon Chambers

The muon sub-detector system of the CMS experiment [157] is the outer most part of the detector. It is situated completely outside of the CMS magnet. The muon chambers are interleaved with the steel return yoke. The whole system is split up into five wheels, making up the barrel region, and two endcap regions. Each wheel of the barrel region consists of four layers of rectangular Drift Tube (DT) chambers, following the same 12-fold segmentation in ϕ as the steel return yoke. The barrel region covers the range of $|\eta| < 1.2$. More muons with a larger background flux are expected in the endcap region compared to the barrel region. In this regard, Cathode Strip Chambers (CSC) are used since they have a fast response time, a high radiation tolerance and allow for a fine segmentation. The endcap regions also consist of four layers, while the pseudorapidity range is equal to $|\eta| \in [0.9, 2.4]$. A third type of detector, i.e. Resistive Plate Chambers (RPC), are intersected in both the barrel and the endcap region. The RPC are to aid with the measurement of the track path, allowing for a more precise measurement of the deflection of the muons due to the magnetic field. A quarter slice of the muon system is shown in Fig. 4.7.

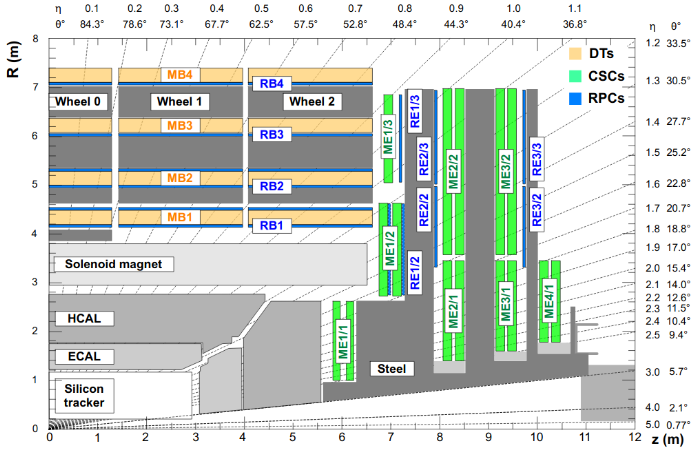


Figure 4.7: An overview (quarter slice) of the CMS muon system. The position of all sub-detectors is shown. The four layers of DT chambers are labeled by MB1 through MB4, where MB stands for muon barrel. The RPC in the barrel region are indicated with RB along with their layer number, while they are denoted as RE in the endcaps. The label ME or muon endcap is used for the CSC detectors. Taken from [158].

The information harnessed by the muon chambers and the tracker are combined in the muon reconstruction [159]. The reconstruction efficiency for muons with a p_{T} of at least a couple of GeV is greater than 95%, while the rate of misidentifying a hadron for a muon is well below 1%. Depending on the detector region, the p_{T} resolution varies between 1% and 6% for muons with a transverse momentum smaller than 100 GeV. For high-energetic muons in the central region with a p_{T} smaller than 1 TeV, the resolution is still better than 10%.

4.2.6 The Trigger System

The trigger system is used to filter out rare or interesting events and to suppress background events. At the LHC, proton bunches cross each other at the interaction point of the CMS detector at a rate of 40 MHz, with a peak instantaneous luminosity of $L \sim 10^{34}$ cm$^{-2} \cdot$ s^{-1} for the second period of data taking (Run II). The CMS experiment uses a two-step trigger system in order to reduce the initial rate of events of 40 MHz to 100 Hz. A schematic overview of the trigger system is shown in Fig. 4.8.

The first of the two steps is the Level 1 (LV1) Trigger [160, 161], which is used to reduce the rate of events to a maximum of 100 kHz, with a maximum latency time of 3.8 μs. The LV1 Trigger uses custom electronics and a coarser segmentation of the calorimeters and the muon chambers to lessen the amount of channels. The LV1 trigger system uses input from the calorimeters and the muon chambers, and bases its decision

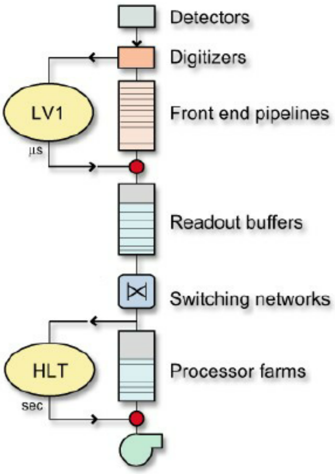


Figure 4.8: A schematic sketch of the two-level trigger system used at the CMS experiment. The two step trigger system is used to reduce the rate of 40 MHz at the interaction point down to 100 Hz.

on the presence of particle candidates, e.g. electrons, muons, jets, . . ., above a set energy or transverse momentum threshold.

If an event passes the LV1 Trigger, then the High-Level Trigger (HLT) [162, 163] will reduce the event rate further down to 100 Hz. The HLT uses a processor farm to perform a further selection for which the maximal processing time per event is 320 ms. The selection process has implemented a light version of the offline reconstruction, which has been optimized for online selection. All events selected by the HLT are stored by the Data Acquisition (DAQ)

The data that has been selected by the trigger system undergoes online and offline Data Quality Monitoring (DQM) [164]. The online DQM happens during the detector operation Histograms are filled with data from the events and are monitored in order to detect possible errors coming from the detector hardware of the reconstructions during the run itself. While offline DQM goes into more detail in order to look for possible errors. When this process is completed, the final data set files are produced along with the JavaScript Object Notation (JSON) files. The JSON files hold all information on which runs and luminosity sections contain good and certified data. The final data samples are ready to be used in physics analyses.

Jet Triggers

Jets are often central objects in QCD analyses but also play an important role in searches for new physics. Therefore CMS has dedicated jet triggers [165]. In this work, single jet and dijet triggers will both be used. The former plays a central role in the physics analysis, while the latter triggers are used in the determination of the Jet Energy Correction (JEC) factors in Sec. 6.

The LV1 jet triggers use energy sums of the ECAL and HCAL in the central region and of HF in the forward region. In the central region the energy sums are computed in a 4 by 4 Trigger Tower (TT) matrix located around a local maximum called the jet *seed*. While in the forward region 4 or 6 TTs are used. Each TT consists of multiple adjacent detector cells that have been combined into coarser detector segments. Whenever a deposit of transverse energy in the calorimeters exceeds the seed threshold, the event is selected by the LV1 Trigger system and passed on to the HLT. Next, the anti-k$_T$ clustering algorithm [104] with a jet cone size R equal to 0.4 reconstructs the jets. If the clustered jet passes the HLT, then the event is stored by DAQ and processed by the DQM. As the whole process takes a rather significant amount of CPU resources and time, while the rate of jet events is quite high, the jet triggers are assigned a LV1 and HLT prescale. If a LV1 trigger with a LV1 prescale of N is fired a multitude of times, only one per N events will be probabilistically chosen and passed on the the HLT. The HLT prescale works in the same way. Generally the prescale is connected to the p_T threshold: a trigger with a high p_T threshold will have a low prescale and vice versa.

Chapter 5

Data Samples

In this chapter, all data samples collected by the CMS detector and used in the analysis will be discussed. The simulation of the CMS detector will also be highlighted since two MC samples have been produced at generator and detector level. Such simulated samples are essential in order to correct data for detector effects because one wishes to obtain detector independent results.

5.1 Data Samples

The CMS detector accumulated data of proton-proton collisions at a centre-of-mass energy of $\sqrt{\mathrm{s}} = 13$ TeV during the Run II program of the LHC in the year 2016. The LHC delivered an integrated luminosity of 40.99 fb^{-1}, while the CMS detector effectively recorded an integrated luminosity of 37.80 fb^{-1} [166]. Where the integrated luminosity is equal the instantaneous luminosity integrated over time or

$$L_{\mathrm{int}} = \int L \; dt \tag{5.1}$$

The integrated luminosity is a measure for the total amount of recorded data. Small sub-samples of the entirety of the available data have been selected, i.e. low pileup runs from the run period Run2016B with low p_{T} collisions where the HF sub-detector of the HCAL was active. Pileup refers to the additional proton-proton collisions, which are either coming from the same bunch crossing (in-time pileup) or from different bunch crossings (out-of-time pileup).

The jets created in a SPS event and in a DPS event exhibit different correlations in certain regions of the DPS-sensitive observables since the jets originating from the two different interactions in a DPS event can be approximated as completely uncorrelated. Jets originating from different interactions in the bunch crossing, i.e. jets coming from pileup interactions, are uncorrelated as well and can possibly mimic a DPS signature in the data. Sub-samples of all the available data with low pileup have been selected as to avoid further necessary pileup corrections and possible fake DPS signatures in the data. Of the low pileup data, events with low p_{T} jets have been selected further. Forward/backward jets combined with small values of p_{T}, allow for the low-x region to be probed, see eq. 3.20, in which the gluon density rises exponentially. The probability of the occurrence of a DPS event in this region of the phase space is expected to rise due to a much denser proton. In general, DPS is expected to play a more significant role in the low pileup, low p_{T} data compared to data containing high p_{T} jets.

From the sub-samples, the data taken in good operation conditions is selected through the use of the JSON file[1] provided by the CMS collaboration. Afterwards, the event information of runs and luminosity sections selected by the JSON file is processed in the `Common FSQ Framework (CFF)`[2], within the CMSSW 8.0.29 framework[3]. All samples are in the `mini Analysis Object Data (miniAOD)` format, from which a so-called *skim* is created, i.e. the information necessary to perform the analysis is extracted from the events and stored separately in order to reduce the total amount of information and to analyze the samples more efficiently. When extracting the *skim* from the totality of data, the global tag `80X_dataRun2_2016SeptRepro_v7` is used. The global tag is a set of database tags which together define the offline conditions data are collected in. In this manner, the conditions database is decoupled from the CMSSW release and different global tags can be used with a given CMSSW release. An overview of the data samples is shown in Tab. 5.1.

Table 5.1: An overview of the data samples used in the analysis, along with the integrated luminosity and the number of raw events, which is the amount of events before the information contained in the JSON file has been taken into account.

Data sample	L_{int} (pb^{-1})	Raw events
HINPFJets/Run2016B-PromptReco-v2/MINIAOD	0.041855118	$6,034,698$
FSQJets/Run2016B-03Feb2017_ver2-v2/MINIAOD	0.041855118	$5,416,877$
L1MinimumBias$X$/Run2016B-PromptReco-v2/MINIAOD, $X \in [0, \ldots, 9]$	0.041855118	$\sim 42,000,000$

5.2 Simulation of the CMS Detector

As mentioned in the discussion on the simulation of hadron collisions, different MC generators will be used to study the data at generator level. However, detector effects need to be corrected for. To this extent, generated MC samples are used to produce MC samples on detector level, i.e. the response of the whole CMS detector is simulated and the effects of all different parts are applied to a generator level MC sample. A subgroup of the CMS collaboration is responsible for the complete mapping and simulation of the detector response. The GEANT4 software package [167, 168] is used. The different steps entailing the simulation of the CMS detector, which happen within the CMSSW framework, are listed below.

---

[1] `/afs/cern.ch/cms/CAF/CMSCOMM/COMM_DQM/certification/Collisions16/13TeV/ReReco/Final-/Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON_LowLowPU.txt`

[2] `https://twiki.cern.ch/twiki/bin/viewauth/CMS/FSQCommonFW`

[3] `https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBook`

- GEN-SIM: the first step produces the events of a MC sample on generator level, where all stable particles are in the final state. The simulated hits of these particles in the detector are produced as well.
- RAW-DIGI: the second step produces the raw detector information by converting the simulated hits into detector response.
- RECO-MINIAOD: the last and third step converts the raw detector information into reconstructed objects. The `miniAOD` data format is used as it is a subset of the `RECO` format while still holding all necessary information whilst being smaller in size.

The subsequent MC samples on detector level can therefore be processed in the same manner as the data samples throughout the analysis.

The CMS collaboration has provided two samples fit for this analysis, i.e. a PYTHIA 8 and a HERWIG++ sample using the `CUETP8M1` and the `CUETHS1` tune, respectively. Both samples have been generated without pileup. The samples, listed in Tab. 5.2, are processed in the `CFF` before they are analyzed with the global tag `80X_mcRun2_asymptotic_2016_TrancheIV_v8`.

Table 5.2: The MC samples used to correct for the detector effects in the analysis, along with their respective cross sections and the raw number of simulated events.

MC sample	σ (pb)	Raw events
/QCD_Pt-15to1000_TuneCUETP8M1_Flat_13TeV_pythia8/RunIISummer16MiniAODv2-NoPU_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v2/MINIAODSIM	1976000000	9,998,200
/QCD_Pt-15to1000_TuneCUETHS1_Flat_13TeV_herwigpp/RunIISummer16MiniAODv2-NoPU_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v2/MINIAODSIM	1369000000	9,993,000

CMS collaboration has provided two samples fit for this analysis, i.e. a PYTHIA 8 and a HERWIG++ sample using the `CUETP8M1` and the `CUETHS1` tune, respectively. Both samples have been generated without pileup. The samples, listed in Tab. 5.2, are processed in the `CFF` before they are analyzed with the global tag

Chapter 6

Event Reconstruction

Before any collected data samples can be analysed, the collisions themselves need to be reconstructed from the detector readout. The reconstruction procedure of the three main physics objects (jets, missing transverse energy and vertices) in this analysis will be discussed. Afterwards, the reconstructed jets need to be corrected for detector effects in order to obtain fully calibrated jets. The calibration is performed by sequentially multiplying the four-momentum of the jet with the Jet Energy Correction (JEC) factors. However, the JEC factors were not readily available for low pileup, low p_{T} data and had to be derived. The determination of the proper JEC factors is fully detailed in the text below and will be the bulk of this chapter.

6.1 Analysis Objects

The Particle-Flow (PF) algorithm [169] aims to reconstruct and identify each individual particle in an event with an optimized combination of information from the various elements of the CMS detector. The energy of photons is obtained from the ECAL measurement. The energy of electrons is determined from a combination of the electron momentum at the primary interaction vertex as determined by the tracker, the energy of the corresponding ECAL cluster and the energy sum of all bremsstrahlung photons spatially compatible with originating from the electron track. The energy of muons is obtained from the curvature of the corresponding track. The energy of charged hadrons is determined from a combination of their momentum measured in the tracker and the matching ECAL and HCAL energy deposits, which are corrected for zero-suppression effects and for the response function of the calorimeters to hadronic showers. Finally, the energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energies.

After the reconstruction and identification of the particles in the event, the information is used to reconstruct different objects that are needed to perform this measurement. The central objects in this work are jets, the missing transverse momentum $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ and vertices.

6.1.1 Jets

After the reconstruction and identification of the particles by the PF algorithm, the infrared and collinear safe anti-k_T algorithm [104, 170] with a distance parameter of 0.4 is used to cluster hadronic jets. The momentum of the jets is determined as the vectorial sum of the momentum of all particles in the jet. Tracks and calorimetric energy depositions originating from pileup can contribute to the momentum. Therefore, the Charged Hadron Subtraction (CHS) is made, removing contributions from particles that have been identified as coming from pileup vertices. The collection of reconstructed jets obtained in this manner, is be labeled as the `AK4PFCHS` jets.

Further corrections need to be made since the jets are not yet fully calibrated. A schematic view of the different corrections is given in Fig. 6.1. Generally, the JEC are provided by the `JetMet` subgroup of the CMS collaboration, however, the corrections were not readily available for low pileup and low p_T data from the year 2016, and will be calculated in what follows.

6.1.2 Missing Transverse Momenta

The raw Missing Transverse Momentum (MET) vector $\vec{p}^{\,\mathrm{miss}}_{\mathrm{T,RAW}}$ is computed as the negative vector sum of the transverse momenta of all the PF candidates in an event [171]. The MET is needed in the calculation of the L2L3 residual corrections, as detailed in Sec. 6.2.2. The MC-truth corrections, derived in Sec. 6.2.1, will have been applied to the jets at this point in the analysis. To take the corrections into account, the MET is rewritten as follows

$$\begin{aligned}
\vec{p}^{\,\mathrm{miss}}_{\mathrm{T,RAW}} &= -\sum_{i\in\mathrm{all}} \vec{p}_{\mathrm{T},i} && (6.1)\\
&= -\sum_{i\in\mathrm{jets}} \vec{p}_{\mathrm{T},i} - \sum_{i\in\mathrm{uncl.}} \vec{p}_{\mathrm{T},i} && (6.2)\\
&= -\sum_{\mathrm{jet}} \vec{p}^{\,\mathrm{raw}}_{\mathrm{T,jet}} - \sum_{i\in\mathrm{uncl.}} \vec{p}_{\mathrm{T},i} && (6.3)
\end{aligned}$$

First, the distinction between particles clustered in jets and unclustered particles is made. Next, the summation over the transverse momentum of all particles in jets is replaced by the summation over the uncorrected transverse momentum of all jets. Next, the effect of the MC-truth corrections has to be included through the use of the so-called type-I correction factor given below.

$$\vec{C}^{\,\mathrm{Type-I}}_{\mathrm{T}} = \sum_{\mathrm{jet}} \vec{p}^{\,\mathrm{raw}}_{\mathrm{T,jet}} - \sum_{\mathrm{jet}} \vec{p}^{\,\mathrm{L123}}_{\mathrm{T,jet}} \tag{6.4}$$

All of the MC-truth corrections together are often denoted as L123, where the abbreviated notation refers to the L1 offset and L2L3 MC truth corrections. The correction factor is the difference between the total uncorrected transverse momentum of all jets with the total transverse momentum of all jets where the MC-truth corrections have been applied. Adding the type-I correction factor to the raw MET, yields the type-I corrected MET or

$$\vec{p}_{\mathrm{T,Type-I}}^{\mathrm{miss}} = \vec{p}_{\mathrm{T,RAW}}^{\mathrm{miss}} + \vec{C}_{\mathrm{T}}^{\mathrm{Type-I}} \tag{6.5}$$

$$= -\sum_{\mathrm{jet}} \vec{p}_{\mathrm{T,jet}}^{\mathrm{L123}} - \sum_{i\in\mathrm{uncl.}} \vec{p}_{\mathrm{T},i} \tag{6.6}$$

6.1.3 Vertices

Vertex reconstruction is performed in three steps. First, tracks are selected. afterwards, the tracks that seem to originate from the same interaction vertex are clustered together. As the last step, a fit is performed in order to determine the position of each vertex associated with the tracks. A more detailed description can be found in [149]. The candidate vertex with the largest value of summed physics-object p_{T} is taken as the primary interaction vertex. The physics objects are the jets and the associated missing transverse momentum.

6.2 Jet Energy Corrections

The `AK4PFCHS` jets need to be calibrated to compensate for detector effects. The CMS collaboration has adopted a factorized approach. The detector effects are corrected for by different levels of JEC applied sequentially. The effect of each of the different JEC is a scaling of the four momentum of the jet, where the scaling factor is dependent on the jet quantities such as p_{T}, η and pileup. Once the correction has been applied, the resulting four-momentum is taken as the input for the next correction. The different corrections and their order are schematically depicted in Fig. 6.1. The flavor correction in the figure is of no importance in this study. Only jets originating from an up, down or strange quark, the so-called light jets, are considered in this analysis. Events with heavy jets, i.e. jets initiated by a charm or bottom quark, are not excluded, however, their contributions are negligible for the interaction process of this analysis.

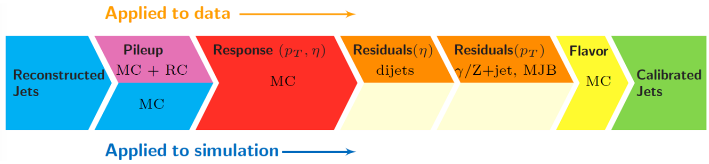


Figure 6.1: Diagrammatic representation of the application of JEC for data and MC. The flavor corrections are not applied in this analysis. Only light jets initiated by an up, down or strange quark are considered. Taken from [172].

The jets will undergo the so-called *jet ID* selection criteria before any of the corrections are applied. The *jet ID* criteria are based on the composition of the reconstructed jets. In Tab. 6.1 and Tab. 6.2, the *jet ID* criteria are shown for jets in the central and forward regions of the detector, respectively. The quality criteria are applied in order to reject fake jets, which can arise from detector noise or wrongly reconstructed particles. All values are provided by the CMS collaboration for the data of Run2016B[1]. In this analysis it was opted to use the tight ID throughout.

Starting on the left side of Fig. 6.1, the first set of corrections to be applied to the jets passing the *jet ID* criteria are meant to correct for additional energy coming from pileup events. The corrections are generally referred to as the L1 offset corrections and are set to one in this analysis for two reasons. First, the data are low pileup data, while in Sec. 7 it will be highlighted that events with only one proper vertex will be used. Second, both detector level MC samples have been generated without any pileup.

The L2L3 MC-truth corrections aim to correct the non-uniformity of the jet response as a function of η and p_{T} by comparing the transverse momentum on generator and detector level. The PYTHIA 8 sample using the `CUETP8M1` tune will be used in this regard. The derivation of the corrections is outlined in Sec. 6.2.1

The third and last set of corrections needed in this analysis are the L2L3 residual corrections. The simulation of the detector level MC samples is not perfect, therefore, small residual differences remain between the MC and data samples. The residual corrections are meant to compensate for the remaining differences and are discussed in detail in Sec. 6.2.2.

Table 6.1: *Jet ID* criteria for the central region of the detector.

PF Jet ID	Loose	Tight	TightLepVeto
jet ID criteria for $\|\eta\| \leq 2.7$			
Neutral Hadron Fraction	< 0.99	< 0.90	< 0.90
Neutral EM Fraction	< 0.99	< 0.90	< 0.90
Number of Constituents	> 1	> 1	> 1
Muon Fraction	-	-	< 0.8
Additional *jet ID* criteria for $\|\eta\| \leq 2.4$			
Charge Hadron Fraction	> 0	> 0	> 0
Charged Multiplicity	> 0	> 0	> 0
Charged EM Fraction	< 0.99	< 0.99	< 0.90

Table 6.2: The *jet ID* criteria for the forward region of the detector.

PF Jet ID	Loose	Tight
Jet ID criteria for $2.7 \leq \|\eta\| \leq 3.0$		
Neutral Hadron Fraction	> 0.01	> 0.01
Neutral EM Fraction	< 0.98	< 0.98
Number of Constituents	> 2	> 2
Jet ID criteria for $\|\eta\| \geq 3.0$		
Neutral Hadron Fraction	> 0.01	> 0.01
Neutral EM Fraction	< 0.98	< 0.98
Number of Constituents	> 2	> 2

---

[1] `https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2016`

6.2.1 L2L3 MC-truth Corrections

For the determination of the L2L3 MC-truth corrections, the `JetMet` group has provided a framework, i.e. the `JetMetAnalysis`/`JetUtilities` and `JetMetAnalysis`/`JetUtilities` packages[2]. The L2L3 MC-truth corrections are meant to correct the non-uniformity of the jet response, defined as

$$\text{Response}(p_{\text{T}}, \eta) = \frac{p_{\text{T}}^{\text{reco}}}{p_{\text{T}}^{\text{gen}}} \tag{6.7}$$

Where the response function is taken in bins of p_{T} and η. The response is calculated for all matching detector and generator level jets, resulting in peaked distributions for each bin in p_{T} and η. The distributions are approximated by and fitted with a Gaussian function. For all bins, the median and the error on the median of the response distributions are taken as the correction and error on the correction, respectively. The whole calculation along with the production of the JEC files is handled by the provided framework. However, some alterations to the `JetMetAnalysis`/`JetUtilities` package have been made

- In the file `L2Creator.cc` the minimum number of entries for the response distributions in line 235 has been changed from 4 to 100 in order to combat a possible lack in statistics and the resulting bias thereof.
- In the file `HistogramUtitlities.cc` the lines 125-137 have been removed, where the calculation of the error on the median has been changed to

  $$\sigma_{median} = 1.253 \cdot \frac{\sigma}{\sqrt{N}} \tag{6.8}$$

  Where σ is the standard deviation of the Gaussian distribution and N is the amount of entries. The definition is valid as the distribution is approximated with a Gaussian distribution and enough statistics are present in each bin, in this case at least 100 events.

After the implementation of these changes the total L2L3 MC-truth corrections have been calculated using the PYTHIA 8 sample using the `CUETP8M1` tune. The resulting corrections are shown as a function of η for different values of p_{T} in Fig. 6.2. A complete overview of the corrections in the (p_{T}, η)-plane can be found in Fig. 6.3. Note that not for all values of p_{T} and η corrections have been calculated due to a lack in statistics. In this analysis, almost no jets exceed 800 GeV and very few jets with a high p_{T} lie out of the central area of $|\eta| < 2.5$. However, if and when a jet with a certain p_{T} lies out of the rapidity region for which corrections are available, then it receives the correction for a jet with the same p_{T} but with the closest value of η for which a correction has been calculated.

Closure Test

A closure test is performed as a cross check. The newly obtained corrections are applied to the same sample used to derive them. Afterwards, the whole calculation is repeated.

---

[2] `https://github.com/cms-jet/JetMETAnalysis/tree/master/JetAnalyzers`

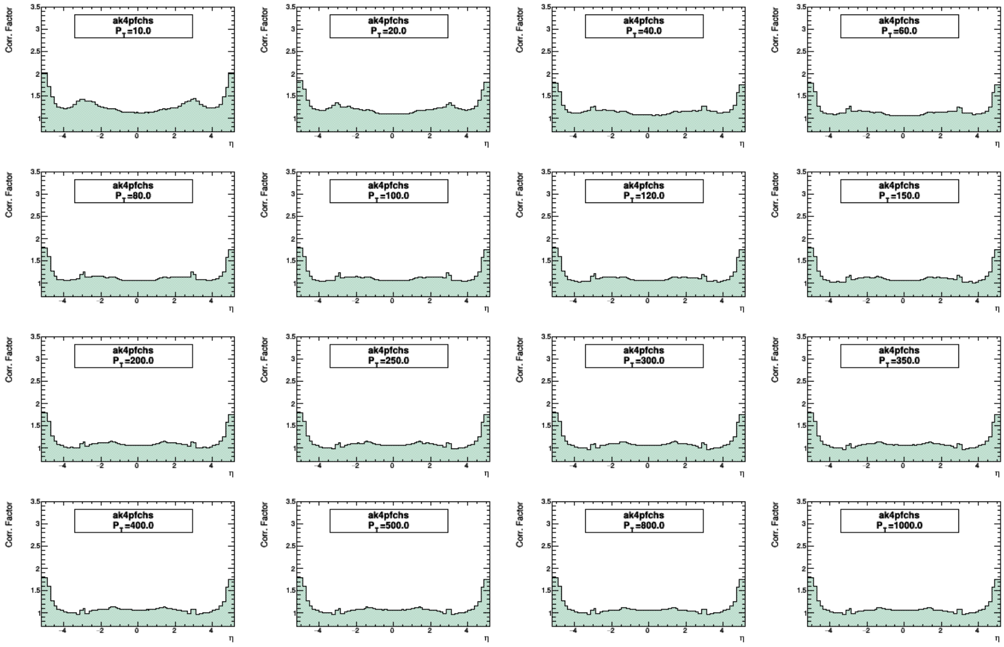


Figure 6.2: L2L3 MC-truth correction factors as a function of the η for different values of p_{T}.

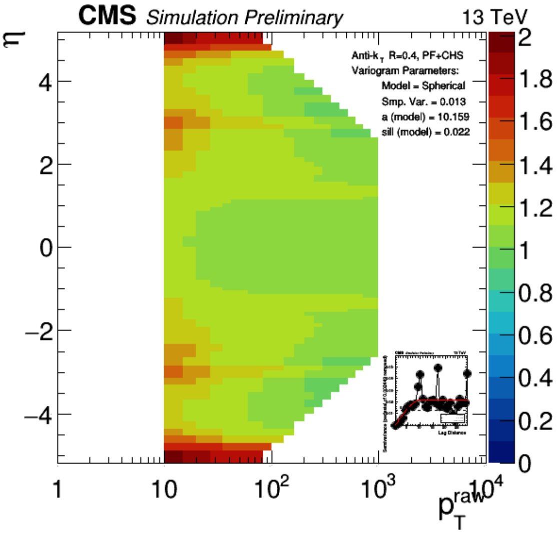


Figure 6.3: Overview of the L2L3 MC-truth corrections in the (p_{T}, η)-plane.

If the corrections are precise, then the average of the response distributions (6.7) should be approximately equal to one. As a rule of thumb, closure of the L2L3 MC-truth corrections is achieved whenever the average does not deviate more than 0.01 from unity. The averages are plotted as a function of the pseudorapidity of the reconstructed jets η^{reco} and per bin in the transverse momentum at generator level $p_{\text{T}}^{\text{ptcl}}$. The results are shown in the appendices in Fig. C.1, Fig. C.2, Fig. C.3 and Fig. C.4. Closure is achieved throughout the plots except for a few forward/backward rapidity bins at high p_{T}, which is due to a lack of statistics in these bins. Therefore, it can be concluded that the newly derived corrections display the proper behavior and have the desired effect on the response functions.

The corrections have been approved by the `JetMet` group and have been published in their online JEC database[3]. From hence on forth these corrections will be applied to all data and MC samples, that contain events at detector level.

6.2.2 L2L3 Residual Corrections

The remaining differences between the jet response in data and MC are corrected for by the L2L3 residual corrections. Dijet and multijet events are used in the derivation following methods outlined in [172–175]. One of the jets, also called the tag jet, is required to be in the well calibrated barrel region $|\eta| \leq 1.3$. The so-called probe jet is used to calculate the L2L3 residual corrections relative to the barrel jet. The pseudorapidity window of the probe jet is limited to $|\eta| < 4.7$. Therefore, the corrections will be relative to the barrel region as a whole. The corrections are derived as function of η and $p_{\text{T,Ave}}$, where $p_{\text{T,Ave}}$ is given by

$$p_{\text{T,Ave}} = \frac{1}{2}\left(p_{\text{T,tag}} + p_{\text{T,probe}}\right) \tag{6.9}$$

In the ideal dijet topology, the tag and probe jets are back-to-back in the azimuthal plane. However, the momentum balance of the dijet can be affected by additionally radiated jets. The presence of the jets will be corrected for by performing an extrapolation of the parameter $\alpha \to 0$, where α is defined below.

$$\alpha = \frac{p_{\text{T,3rd}}}{p_{\text{T,Ave}}} \tag{6.10}$$

The perfect dijet topology is approximated since the transverse momentum of the third and lower leading jets is negligible when $\alpha \to 0$. The derivation of the L2L3 residual corrections is done for the `AK4PFCHS` jet collection.

Data Sample and Event Selection

For the derivation of the L2L3 residual corrections, the `FSQJets` data sample and the PYTHIA 8 MC sample using the `CUETP8M1` tune have been used. The `FSQJets` data sample is preferred over the `HINPFJets` data sample. The reason for this choice is threefold. First, the `PFJet` triggers and the `DiPFJet` triggers yield more statistics in the `FSQJets`

---

[3] `https://github.com/cms-jet/JECDatabase/tree/master/textFiles/Summer16_lowPU_V1_MC`

sample compared to the `HINPFJets` sample. Second, the `DiPFJetAve` triggers, especially designed for studies of the L2L3 residual corrections, have non-zero rates in the `FSQJets` sample, which is not the case for the `HINPFJets` data sample. Third, the `PFJet` and `DiPFJet` triggers have lower p_T thresholds. A larger region of the phase space at low p_T can be probed, compared to the `AK4PFJet` triggers used in the four jet analysis.

The various triggers become efficient at different values of p_T and have different prescales, see Sec. 9.1 for additional information. The phase space is divided in disjunctive areas as a function of the triggers to avoid selecting an event more than once. Generally, the prescale of a trigger is higher when the p_T threshold of the trigger is lower, making triggers with a high threshold more effective when filtering out events with high-p_T jets. The exact division of the phase space as a function of the p_T of the two leading jets is outlined below.

- The `PFJet` and `DiPFJet` triggers will be used whenever $|\eta_\mathrm{probe}| < 2.7$. The exact division of the phase space as a function of the p_T of the two leading jets is given in Fig. 6.4.
- If the probe jet lies in the forward region, i.e. $|\eta_\mathrm{probe}| > 2.7$, the `DiPFJetAve` triggers will be used. The phase space is divided as a function of $p_\mathrm{T,Ave}$ in the following way:
  - ⋆ `HLT_DiPFJetAve_15_HFJEC_v1`: 15 GeV $\leq p_\mathrm{T,Ave} \leq$ 35 GeV
  - ⋆ `HLT_DiPFJetAve_25_HFJEC_v1`: 35 GeV $\leq p_\mathrm{T,Ave} \leq$ 55 GeV
  - ⋆ `HLT_DiPFJetAve_35_HFJEC_v1`: $p_\mathrm{T,Ave} \geq$ 55 GeV

In both cases the tag jet is situated in the barrel region ($|\eta| < 1.3$). A correction due to the trigger efficiency, see Sec. 9.1, is applied wherever necessary. Further selection criteria are applied to all selected events, they are given below.

- At least one good vertex is present in the events:
  - ⋆ The Number of Degrees of Freedom (NDF) of a vertex is strictly larger than four.
  - ⋆ Spatial constraints on the vertex position are $|z| \leq 24$ cm and $\rho \leq 2$ cm, where ρ is the radial coordinate.
  - ⋆ Fake vertices are discarded. They can be produced by secondary interactions with the detector material, by weak decays or by tracks coming from the beam spot.
- At least two jets passing the *jet ID* selection criteria are present.
- At least one of the two jets is situated in the barrel region $|\eta| \leq 1.3$.
- The jets are approximately in a back-to-back configuration or $|\phi_1 - \phi_2| \geq 2.7$.
- A binning in the maximum allowed value of the parameter α_max is introduced, where $\alpha \leq \alpha_\mathrm{max}$ and $\alpha_\mathrm{max} \in \{0.05, 0.10, \ldots, 0.40, 0.45\}$.

The last two selection criteria were introduced as a perfect dijet topology is rarely observed due to ISR and FSR.

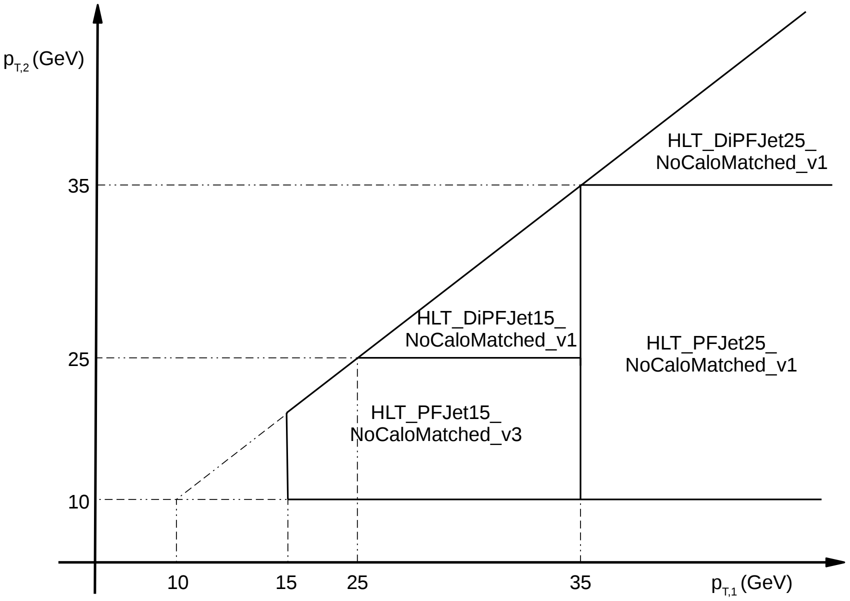


Figure 6.4: The division of the phase space as a function of the p_T of the two leading jets. The division is made so that the triggers do not overlap and events can not be selected twice.

Methodology

Two different methods have been used and are compared to one another in the calculation of the L2L3 residual corrections. The first method is the p_T-balance method, which has been used for the CMS jet energy corrections in the past [172–174]. The p_T-balance method evaluates the asymmetry variable A and the corresponding jet energy response $R^{p_T}_{\rm rel}$, defined as

$$R^{p_{\rm T}}_{\rm rel} = \frac{1+\langle A\rangle}{1-\langle A\rangle}, \quad \text{where} \tag{6.11}$$

$$A = \frac{p_{\rm T,probe} - p_{\rm T,tag}}{2p_{\rm T,Ave}} \tag{6.12}$$

The second method is the Missing transverse momentum Projection Fraction (MPF) method [172, 173, 175], where jet energy response function is defined as

$$R^{\rm MPF}_{\rm rel} = \frac{1+\langle B\rangle}{1-\langle B\rangle}, \quad \text{where} \tag{6.13}$$

$$B = \frac{\vec{p}^{\,\rm miss}_{\rm T,Type-I} \cdot (\vec{p}_{\rm T,tag}/p_{\rm T,tag})}{2p_{\rm T,Ave}} \tag{6.14}$$

The $\vec{p}_{\mathrm{T,Type-I}}^{\,\mathrm{miss}}$ is the type-I corrected MET, described in Sec. 6.1.2, and is also referred to as the MET corrected in the `(L123(L23)-L1RC)` scheme[4]. The MPF method is expected to be less sensitive to additional jet activity compared to the p_{T}-balance method. The full hadronic recoil of the event is taken into account through the use of $\vec{p}_{\mathrm{T,Type-I}}^{\,\mathrm{miss}}$ in the response function, whereas the p_{T}-balance method only considers the two leading jets as a dijet system.

The p_{T}-balance and MPF response will both be calculated as a function of p_{T} in each $|\eta_{\mathrm{probe}}|$ bin for a value of $\alpha_{\mathrm{max}} = 0.3$ for the data and MC samples. The residual differences between the data and MC samples are studied as a function of the ratio of the response functions given below.

$$\left(\frac{R_{\mathrm{rel}}^{\mathrm{MC}}}{R_{\mathrm{rel}}^{\mathrm{Data}}}\right)_{\alpha\leq 0.3} \tag{6.15}$$

Through fitting a linear logarithmic function $p_0 + p_1 \log(p_{\mathrm{T}})$, the p_{T} dependence of the response functions is extracted. However, both responses need to be corrected for the additional jet activity. In this regard, the k_{FSR} correction factor needs to be obtained. The first step is to determine the factors $k_{\mathrm{FSR}}(\alpha_{\mathrm{max}}, |\eta_{\mathrm{probe}}|)$ defined below.

$$k_{\mathrm{FSR}}(\alpha_{\mathrm{max}}, |\eta_{\mathrm{probe}}|) = \left(\frac{R_{\mathrm{rel}}^{\mathrm{MC}}}{R_{\mathrm{rel}}^{\mathrm{Data}}}\right)_{\alpha\leq\alpha_{\mathrm{max}}} \Bigg/ \left(\frac{R_{\mathrm{rel}}^{\mathrm{MC}}}{R_{\mathrm{rel}}^{\mathrm{Data}}}\right)_{\alpha\leq 0.3} \tag{6.16}$$

$$\text{with} \quad \alpha_{\mathrm{max}} \in [0.05, 0.10, \ldots, 0.45] \tag{6.17}$$

The $k_{\mathrm{FSR}}(\alpha_{\mathrm{max}}, |\eta_{\mathrm{probe}}|)$ factors are determined as a function of α_{max} in each $|\eta_{\mathrm{probe}}|$ and $p_{\mathrm{T,Ave}}$ bin. Next, the weighted mean over $p_{\mathrm{T,Ave}}$ is taken for each value of α_{max} and $|\eta_{\mathrm{probe}}|$. The final correction factor k_{FSR} for each η_{probe} bin is obtained by extrapolating the fitted functions to $\alpha_{\mathrm{max}} = 0$.

$$k_{\mathrm{FSR}}(|\eta_{\mathrm{probe}}|) = \lim_{\alpha_{\mathrm{max}}\to 0} k_{\mathrm{FSR}}(\alpha_{\mathrm{max}}, |\eta_{\mathrm{probe}}|) \tag{6.18}$$

The $k_{\mathrm{FSR}}(|\eta_{\mathrm{probe}}|)$ factor is determined as an average correction factor in $p_{\mathrm{T,Ave}}$. The L2L3 residual corrections relative to the barrel region can then be written as the product of the $k_{\mathrm{FSR}}(|\eta_{\mathrm{probe}}|)$ factor multiplied with the p_{T} dependent linear logarithmic function extracted from the ratio in eq. (6.15).

$$\mathrm{L2L3Res}(|\eta_{\mathrm{probe}}|) = k_{\mathrm{FSR}}(|\eta_{\mathrm{probe}}|) \cdot \left(\frac{R_{\mathrm{rel}}^{\mathrm{MC}}}{R_{\mathrm{rel}}^{\mathrm{Data}}}\right)_{\alpha\leq 0.3} \tag{6.19}$$

Where the linear logarithmic fit will be used to include the p_{T} dependence of the L2L3 residual corrections. Dividing by the average correction over the barrel region will result in the absolute L2L3 residual corrections for the `AK4PFCHS` jet collection.

---

[4] `https://twiki.cern.ch/twiki/bin/viewauth/CMS/METType1Type2Formulae#3_The_Type_I_correction`

Calculation of the L2L3 Residual Corrections

To start the calculation of the L2L3 residual corrections, the variables A and B have been measured in bins of $p_{\mathrm{T,Ave}}$ and $\alpha_{\max}$. Fig. C.5, Fig. C.6, Fig. C.7 and Fig. C.8 in the appendices show the resulting distributions for the `FSQJets` and the PYTHIA 8 samples, where $\alpha_{\max} = 0.3$ has been used. In comparing the two variables, one observes that the distributions obtained by the MPF method are narrower than the distributions for the p_{T}-balance method.

From the distributions of A and B, the responses (6.11) and (6.13) can be calculated. Next, the ratio of the responses of the MC and data samples is taken. Fig. C.9 and Fig. C.10 in the appendices show the results for all $|\eta_{\mathrm{probe}}|$ bins for both the p_{T}-balance and MPF method, respectively. The η-dependent corrections are studied in bins of $p_{\mathrm{T,Ave}}$ and fitted by a linear logarithmic function in order to extract the p_{T} dependence. Generally, the results obtained through the MPF method have smaller statistical errors compared to the ratios from the p_{T}-balance method. The observation is directly linked to the observed difference in width of the distributions of the variables A and B. The size of the statistical error is limited due to the amount of selected events of the data sample. The amount of statistics of the PYTHIA 8 sample greatly surpasses that of the `FSQJets` data sample. In general the p_{T} dependence becomes stronger for the more forward bins in $|\eta_{\mathrm{probe}}|$ for both methods. Fig. 6.5 shows the same results for the response ratios but as a function of $|\eta_{\mathrm{probe}}|$ for different values of $p_{\mathrm{T,Ave}}$ for the p_{T}-balance (left) and MPF method (right).

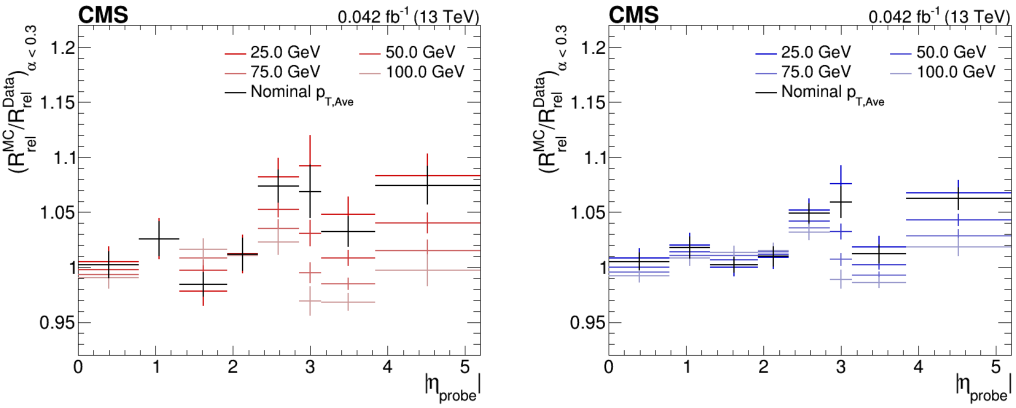


Figure 6.5: The response ratios for the p_{T}-balance (left) and MPF (right) method, shown as a function of $|\eta_{\mathrm{probe}}|$ for different values of $p_{\mathrm{T,Ave}}$. The nominal $p_{\mathrm{T,Ave}}$ is taken as the mean value of the $p_{\mathrm{T,Ave}}$ spectrum in each corresponding rapidity bin.

The next step in the calculation of the L2L3 residual corrections is to correct for the additional jet activity. First, the $k_{\mathrm{FSR}}(\alpha_{\max}, |\eta_{\mathrm{probe}}|)$ factor, given by eq. (6.16), is calculated for each bin in $|\eta_{\mathrm{probe}}|$, $\alpha_{\max}$ and $p_{\mathrm{T,Ave}}$. The results are shown in Fig. C.11 and Fig. C.12 in the appendices for the p_{T}-balance and MPF method, respectively. Second, the weighted average over all $p_{\mathrm{T,Ave}}$ bins per bin in η_{probe} and per $\alpha_{\max}$ is taken. However, before the average curves are fitted, the fitting range has to be determined. As the $k_{\mathrm{FSR}}(|\eta_{\mathrm{probe}}|)$ factor is meant to correct for additionally radiated jets, the correlation between $\langle(p_{\mathrm{T,1st}}^{\mathrm{gen}} - p_{\mathrm{T,2nd}}^{\mathrm{gen}})/p_{\mathrm{T,Ave}}\rangle$ and α_{max} is examined. If the difference in p_{T} of the two leading jets is proportional to the p_{T} of the third leading jet ($p_{\mathrm{T,3rd}}$), then the ratio will be linear in α due to its definition in eq. (6.10). However, if no linear behavior is found, then the fourth, fifth, ... leading jets will have a non-negligible p_{T}, which could affect the extrapolation of $\alpha_{\mathrm{max}} \to 0$ in a negative manner.

The distribution is displayed in Fig. 6.6 and illustrates the correlation between additional radiation in the event and α_{max} [173]. As the discrepancy between the p_{T} of the two leading jets grows, the additional radiation grows linearly with it up to $\alpha_{\mathrm{max}} = 0.32$. Due to limited available statistics, the fitting interval needs to be shortened to $[0.11, 0.32]$. The outliers in the first two bins of α_{max} would yield unstable results with large uncertainties and are, therefore, omitted from the rest of the calculation. The weighted averages of $k_{\mathrm{FSR}}(\alpha_{\mathrm{max}}, |\eta_{\mathrm{probe}}|)$ as a function of α_{max}, are shown in Fig. C.13 and Fig. C.14 in the appendices for the p_{T}-balance and MPF method, respectively. The histograms have been fitted by a linear and a constant fit. When examining the fits compared to the histograms, it becomes apparent why the first two bin have been omitted. The values of the $k_{\mathrm{FSR}}(|\eta_{\mathrm{probe}}|)$ factor for each rapidity bin is then the value of the constant fit or the linear fit evaluated at $\alpha_{\mathrm{max}} = 0$.

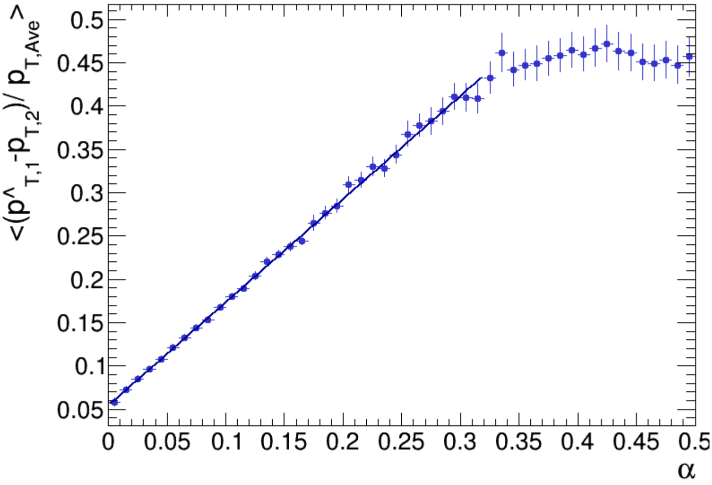


Figure 6.6: The distribution of $\langle(p_{\mathrm{T,1st}}^{\mathrm{gen}} - p_{\mathrm{T,2nd}}^{\mathrm{gen}})/p_{\mathrm{T,Ave}}\rangle$ vs. α_{max} for the PYTHIA 8 sample. The distribution shows linear behavior for $\alpha_{\mathrm{max}} \in [0.00, 0.32]$.

The resulting $k_{\mathrm{FSR}}(|\eta_{\mathrm{probe}}|)$ factors have been bundled in Fig. 6.7, where the left and right panel display the results for the p_{T}-balance and MPF method, respectively. The results of the linear fit for the p_{T}-balance method show a notable difference in the forward region compared to the constant fit. While the MPF method shows little to no significant differences between the results from both fits. For both methods the constant fit yields smaller statistical errors. Since the uncorrected ratios in Fig. 6.5 already yield large errors, it is opted to continue with the constant fit for both methods. If more statistics were available, a more precise linear fit could have been performed.

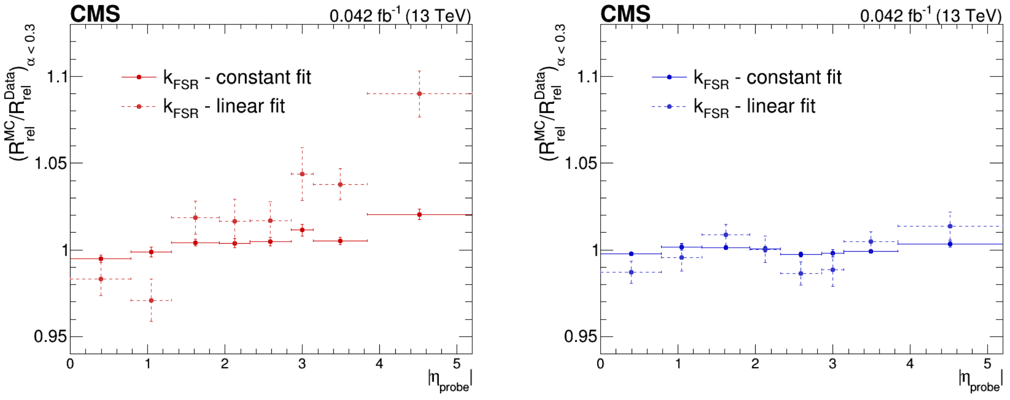


Figure 6.7: The factor $k_{\mathrm{FSR}}(|\eta_{\mathrm{probe}}|)$ for the p_{T}-balance method (left) and MPF (right) method as a function of $|\eta_{\mathrm{probe}}|$. The results from both the constant and linear fit are shown.

The L2L3 residual corrections relative to the barrel region are obtained by multiplying the value of the linear logarithmic fit evaluated at the centre of the $|\eta_{\mathrm{probe}}|$ bins with the corresponding value of the $k_{\mathrm{FSR}}(|\eta_{\mathrm{probe}}|)$ factor for both methods, respectively. The final L2L3 residual corrections are obtained after the normalization to the barrel region. The relative L2L3 residual corrections are divided by the average over the first two bins. Fig. 6.8 shows the final L2L3 residual corrections as a function of $|\eta_{\mathrm{probe}}|$ for different values of $p_{\mathrm{T,Ave}}$ for both methods, respectively.

The final corrections are acquired after multiple iterations where the binning in p_{T} and $|\eta_{\mathrm{probe}}|$ was adapted in each step with the intention of reducing the errors on the corrections, while simultaneously trying to keep the $p_{\mathrm{T,Ave}}$ sensitivity. The re-binning in $|\eta_{\mathrm{probe}}|$ was done in such a manner that the rapidity intervals consist of detector parts that are expected to show similar behavior, while the binning in p_{T} was largely determined by the limited available statistics.

Closure of the L2L3 Residual Corrections

For the L2L3 residual corrections, the closure test is performed in the same manner as it was done for the L2L3 MC-truth corrections. The corrections are applied to the data sample used to derive the corrections, after which the whole analysis is repeated in exactly the same manner. The closure plots for the p_{T}-balance and MPF method are shown in Fig. 6.9.

The derived L2L3 residual corrections do not exhibit closure. The spread of the closure test for different values of $p_{\mathrm{T,Ave}}$ for the MPF method is smaller compared the spread for the closure test for the p_{T}-balance method in general. None of the two sets of corrections show full closure, however, the closure plot for the MPF method shows a more stable behavior in combination with smaller errors. Therefore, the L2L3 residual

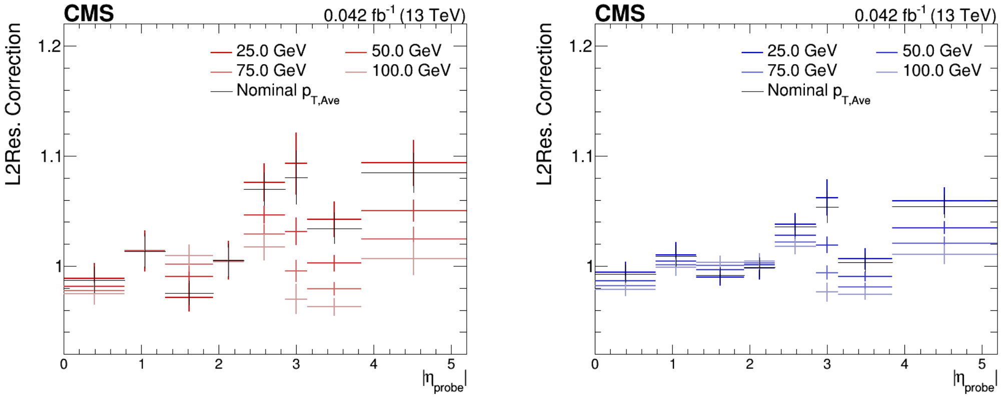


Figure 6.8: The L2L3 residual corrections for the p_{T}-balance (left) and MPF (right) method shown as a function of $|\eta_{\mathrm{probe}}|$ for different values of $p_{\mathrm{T,Ave}}$. The nominal $p_{\mathrm{T,Ave}}$ is taken as the mean value of the $p_{\mathrm{T,Ave}}$ spectrum in each corresponding rapidity bin.

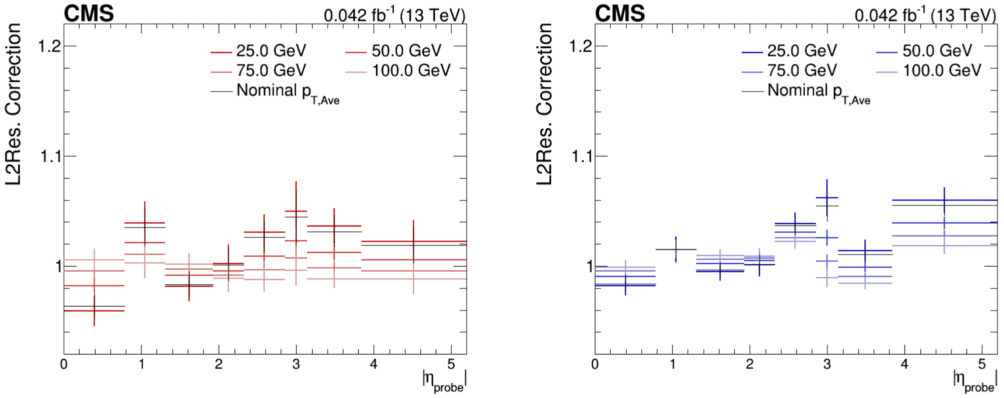


Figure 6.9: The closure plots of the L2L3 residual corrections for the p_{T}-balance (left) and MPF (right) method. The nominal $p_{\mathrm{T,Ave}}$ is taken as the mean value of the $p_{\mathrm{T,Ave}}$ spectrum in each corresponding rapidity bin.

corrections that were obtained by the MPF method will be applied to the data in this analysis.

Ideally, the re-derived corrections would be equal to unity within their respective uncertainty. However, due to only a small amount of data being available in the low p_{T} and low pileup regime, large bins in $p_{\mathrm{T,Ave}}$ and $|\eta_{\mathrm{probe}}|$ had to be used in order to obtain results with statistical errors that were manageable. Preferably, one would like to use a binning in $|\eta_{\mathrm{probe}}|$ that examines every detector segment individually as in [172]. More statistics could alter the p_{T} dependence of the ratios in Fig. C.9 and Fig. C.10. In regard to the determination of the $k_{\text{FSR}}(|\eta_{\text{probe}}|)$ factor, a smaller fitting range had to be used as well as a constant fit instead of a linear fit. Both can yield meaningful differences in the calculation of the L2L3 residual corrections. The combination of the reliance on fits in the computation of the L2L3 residual corrections and the large statistical errors, result in the non-closure of the corrections.

A consistent way of dealing with the non-closure is to treat it as a systematic uncertainty. A detailed description of the handling of the non-closure can be found in Chap. 11. All results have been discussed with and were approved by the `JetMet` group. The final L2L3 residual corrections, that have been used in this analysis, have been added to the JEC database[5].

---

[5] `https://github.com/cms-jet/JECDatabase/tree/master/textFiles/Summer16_lowPU_V4a_DATA`

Chapter 7

Event Selection

As the jets and events as a whole have been reconstructed and calibrated, the detector level data is ready to be analyzed. The first main step is to filter out the events with the desired four jet topology. In this regard, selection criteria have been put into place, all of which will be discussed below.

The events that make up the data samples, given in Tab. 5.1, have passed an online trigger selection. The `AK4PFJetXX` triggers select events with at least one jet above the p_{T} threshold `XX`. Four different triggers have been used in succession in order to maximize the statistics and to prevent the selection of an event more than once. The phase space has been divided in four disjunctive regions based on the p_{T} of the leading jet, specified below. The efficiency due to detector effects of the different triggers, will be examined in Sec. 9.1.

- `HLT_AK4PFJet30_v3` for 35 GeV $\leq p_{\mathrm{T,1st}} \leq$ 65 GeV
- `HLT_AK4PFJet50_v3` for 65 GeV $\leq p_{\mathrm{T,1st}} \leq$ 100 GeV
- `HLT_AK4PFJet80_v3` for 100 GeV $\leq p_{\mathrm{T,1st}} \leq$ 130 GeV
- `HLT_AK4PFJet100_v3` for $p_{\mathrm{T,1st}} \geq$ 130 GeV

After the online selection, the samples have been subjected to multiple offline selection criteria in order to filter out events with the inclusive four jet topology. First, the jets in each event are required to pass the *Jet ID* criteria before the newly derived JEC are applied, as described in Sec. 6.2. If a jet does not pass the criteria, it is discarded from the analysis. Second, exactly one proper reconstructed primary vertex is required to be present in the event. The average vertex multiplicity per event is ~ 1.3, as we are working with low pileup data. The uncorrected detector level cross section as a function of the vertex multiplicity is shown in Fig. 7.1. A vertex is deemed proper, when the following criteria are met.

- The NDF of a vertex is equal to or larger than 5.
- Spatial constraints on the vertex position are $|z| \leq 24$ cm and $\rho \leq 2$ cm.
- Fake vertices are discarded. They can be produced by secondary interaction with the detector material, by weak decays, or by tracks coming from the beamspot.

The reason for the requirement of exactly one proper vertex is two-fold. On the one hand, no further pileup corrections need to be made to the data. The L1 offset corrections have been set to one, justified by the vertex selection criterion. On the other hand, additional jets not coming from the primary vertex, can survive further kinematic cuts, resulting in a possible bias in DPS-sensitive areas of the observables. The DPS event topology is especially sensitive to jets coming from different interaction vertices. However, the luminosity needs to be corrected due to this selection criterion in order to ensure the proper normalization of the data.

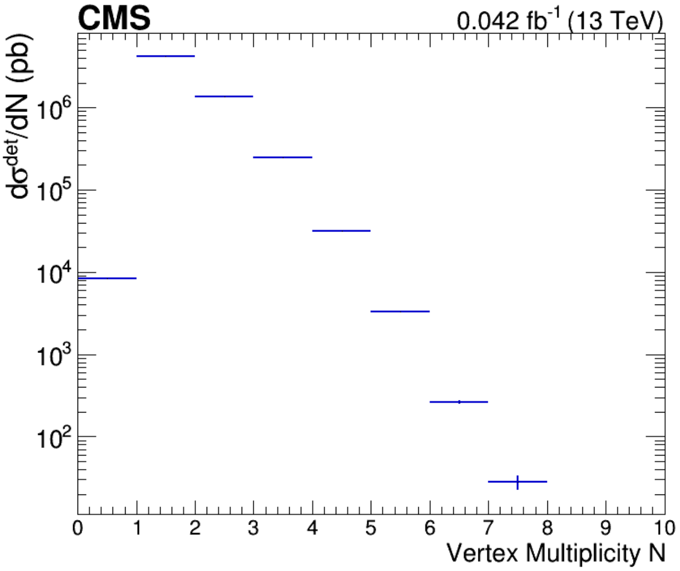


Figure 7.1: The uncorrected detector level cross section as a function of the vertex multiplicity. Roughly 70% of events have exactly one good reconstructed vertex.

The third and final set of cuts focuses on the kinematic properties of the jets. The four leading jets of all events, which have passed all the previous selection criteria, are subjected to the kinematic cuts given below.

- The pseudorapidity of all jets: $|\eta_i| \leq 4.7$
- The p_{T} of the leading jet: $p_{\mathrm{T,1st}} \geq 35$ GeV
- The p_{T} of the second jet: $p_{\mathrm{T,2nd}} \geq 30$ GeV
- The p_{T} of the third jet: $p_{\mathrm{T,4th}} \geq 25$ GeV
- The p_{T} of the fourth jet: $p_{\mathrm{T,4th}} \geq 20$ GeV

The wide pseudorapidity window in combination with low p_{T} jets, allows for the low-x regime to be probed, see eq. 3.20. While a phenomenological study [117] has suggested that asymmetrical cuts could boost the DPS-sensitivity, therefore, it is opted to use these cuts as opposed to symmetrical cuts in contrast to previous studies, e.g. cuts of 50, 50, 20 and 20 GeV on the four leading jets ,respectively, in [26]. A second set of cuts on the jet p_{T} has been applied for the ΔS variable. The second set of cuts are 50, 30, 30 and 30 GeV on the first to last leading jet ,respectively. The exact reasons for these cuts will be detailed in Sec. 8.1 and Sec. 9.2. From hence on forth, the phase space region determined by the softer cuts will be referred to as *region I*, while *region II* indicates the phase space determined by the harder cuts.

Chapter 8

Extraction Strategy of σ_{eff}

In literature, measurements of DPS and the effective cross section σ_{eff} have been performed using two distinct methods.

The first set of techniques can be labeled as the template method, where the DPS contribution is extracted from the measured differential cross section as a function ofone or more DPS-sensitive observables by performing a fit with a signal and a background template. Therefore, the results depend completely on the definition of both templates. The background template is generally constructed from simulation, while for the signal template two approaches have been applied. The signal can be estimated in a data-driven manner or through simulation as well. In the data-driven approaches, events from the process A and B are selected independently from data and are randomly mixed together as to mimic the occurrence of a DPS event. A recent DPS study in the four jet channel performed by the ATLAS collaboration [25] has adopted this approach. When the signal template is is produced through simulation, the preciseness of the predictions of the processes A and B of the used model plays an essential role. For example, W-boson production is a well understood process and has been used in DPS searches in the W + 2-jets channel [13]. Therefore, the signal template was determined from simulation. Let us now return briefly to the background template. As most models include a description of MPI, the background template will have a MPI contribution. MPI is part of the description of the UE, described by the various tunes, and is generally much softer compared to the hard interaction or the two hard scatterings sought after in DPS searches. Therefore, MPI should still be included in the models, as they are less realistic without MPI.

A second method for the extraction of σ_{eff} was introduced in [121]. The differential cross section as a function ofmultiple DPS-sensitive observables was determined as well, but rather than making the distinction between signal and background events the variables are fit inclusively. One requirement is that one should be able to tune the parameters characterizing the UE of the MC event generator. Parameters of the `Tune 4C` were altered in this manner in order to obtain the `CDPSTP8S1-4j` tune. However, it was concluded that the description of the UE and DPS are not compatible. The DPS tunes fail to give a proper description of the UE-sensitive observables and vice versa.

In this analysis, the template method with a data-driven signal template will be used for the extraction of σ_{eff}. A second method will be introduced as well, where the shape of the observable is directly compared in order to estimate the DPS contribution. Both will be detailed below.

8.1 Template Method

The construction of the signal template will be the main challenge and is detailed below. For the background template, the different SPS MC models can be used, i.e. the MC models without an explicit DPS contribution. All the SPS MC models contain a description of MPI, specific to the used UE tune. However, the scale of the MPI is lower compared to the scale of the hard interactions. After the construction of the signal template, the technique used for the template fit and the extraction of the DPS cross section will be highlighted.

8.1.1 Construction of the Signal Template

Let us now come back to the DPS pocket formula given in eq. (3.19). Before the extraction of the DPS cross section $\sigma^{\mathrm{DPS}}_{\mathrm{A,B}}$ can be attempted, the processes A and B need to be identified. In the simplest and ideal topology, the processes A and B would both be dijet production, resulting in a four jet final state with uncorrelated jet pairs, as depicted in Fig. 3.5. However, ISR and FSR, and higher-order interactions can produce final states with more than two jets, effectively dismantling the ideal dijet configuration. Therefore, the processes A and B are both defined as inclusive single jet production. Both processes contribute a number of jets in order to form an inclusive four jet DPS events. If one, two or three of the four leading jets originate from the process A, then the process B has to deliver three, two or one of the four leading jets, respectively.

The threshold of 30 GeV of the lowest single jet trigger imposes further restrictions on the two inclusive single jet processes. The *region I* can not be used since the trigger does not allow for the detection of a single jet with a transverse momentum below 30 GeV. Therefore the *region II* is used, allowing us to define the processes A and B as the inclusive single jet cross sections given below.

$$\sigma_{\mathrm{A}} = \sigma_{\mathrm{jet}}(p_{\mathrm{T}} \geq 50) \tag{8.1}$$

$$\sigma_{\mathrm{B}} = \sigma_{\mathrm{jet}}(p_{\mathrm{T}} \geq 30) \tag{8.2}$$

The cross sections can not be plugged into the DPS pocket formula just yet. Not all combinations of two inclusive single jet events will yield at least four jets in total. In this regard, the "4-jet efficiency" $\epsilon_{4\mathrm{j}}$ is introduced. It represents the fraction of all cases for which the combination of two inclusive single jet events result in at least four jets in the final state, passing all the selection criteria of the four jet analysis.

Events that meet the online and offline selection criteria for the two inclusive single jet processes are selected and stored in two disjunct data sets. Each event with at least one jet with a transverse momentum larger than 50 GeV is randomly mixed with an event where at least one jet has a p_{T} above 30 GeV. To prevent any possible form of double counting, each event with at least one jet with a p_{T} above 30 GeV is used at most once. After such an event is selected, it is omitted from the sample. A veto condition is put in place in order to exclude events where two or more jet axes overlap. The condition is formulated below.

$$R_{ij} = \sqrt{|\phi_i - \phi_j|^2 + |\eta_i - \eta_j|^2} \leq 0.4 \qquad (8.3)$$
$$\text{with } \mathrm{jet}_i \in \mathrm{A},\ \mathrm{jet}_j \in \mathrm{B}$$

The newly constructed pure DPS sample is subjected to the selection criteria of *region II*. A 4-jet efficiency of $\epsilon_{4j} = 0.32441 \pm 0.00053(\text{stat.})$ is found.

Afterwards, a cross check is performed to ensure that the veto condition does not affect the distributions of the observables. The ΔS observable is expected to be one of the more DPS-sensitive observables. Therefore, the detector level ΔS distribution with the veto condition is compared the same distribution where a “naive” re-clustering of the jets has been performed instead of the removal of the event. The re-clustering affects the p_{T}, η and ϕ of the jets. The vectorial combination of the transverse momentum vectors of the overlapping jets is calculated, determining the p_{T} and ϕ of the re-clustered jet. While the pseudorapidity of the overlapping jets is calculated by inverting the definition of the pseudorapidity in eq. (3.13) and by taking the average of the angles θ_i and θ_j. The pseudorapidity of the re-clustered jets is given below.

$$\begin{cases} \theta_i = 2\arctan(e^{-\eta_i}) \\ \theta_j = 2\arctan(e^{-\eta_j}) \end{cases} \Rightarrow \eta_{ij} = -\ln\left(\tan\left(\frac{1}{2}\frac{\theta_i + \theta_j}{2}\right)\right) \qquad (8.4)$$

The detector level ratio of the ΔS distribution with the veto condition in place versus the naive clustering is shown in Fig. 8.1. Both distributions have been normalized to one before the ratio was taken. The relative bin-to-bin differences will be of interest, since the ΔS observable is expected to exhibit a sensitivity to DPS in the tail of the distribution. From the ratio one can conclude that the relative size of the bins does not deviate more than 3%. It is sufficient to continue with the veto condition. A formal re-clustering of the jets is not expected to contribute in a significant way.

One last remark has to be made, before the DPS pocket formula can be rewritten to a workable form. The cross sections of the two processes σ_{A} and σ_{B} are not always distinguishable. The cross section of the process B is split in the two disjunct cross sections σ_{A} and $\sigma_{\mathrm{B}} - \sigma_{\mathrm{A}}$ in order that the symmetry factor can be unambiguously defined. Taking all of the considerations above into account, one obtains the expression for the DPS pocket formula given below.

$$\begin{aligned} \sigma^{DPS}_{\mathrm{A,B}} &= \frac{\epsilon_{4j}}{\sigma_{\mathrm{eff}}}\left(\frac{1}{2}\sigma^2_{\mathrm{A}} + \sigma_{\mathrm{A}} \cdot (\sigma_{\mathrm{B}} - \sigma_{\mathrm{A}})\right) \\ &= \frac{\epsilon_{4j}\sigma_{\mathrm{A}}\sigma_{\mathrm{B}}}{\sigma_{\mathrm{eff}}}\left(1 - \frac{1}{2}\frac{\sigma_{\mathrm{A}}}{\sigma_{\mathrm{B}}}\right) \end{aligned} \qquad (8.5)$$

The cross sections σ_{A} and σ_{B} will be determined from the pseudorapidity spectra of the leading jet for both inclusive single jet processes. The distributions will be corrected to generator level in Chap. 10, for which the proper corresponding MC samples are needed.

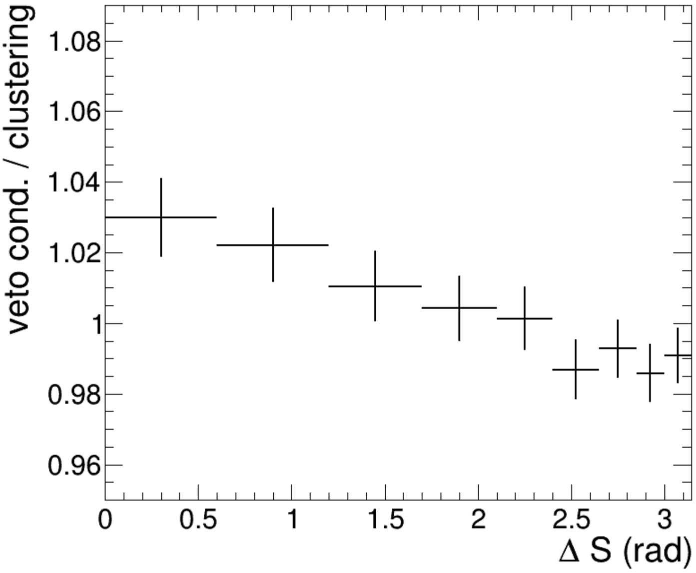


Figure 8.1: The ratio of the detector level ΔS distribution with the veto condition and with the naive re-clustering of the jets. Both distributions have been normalized to one, as the relative bin-to-bin differences are of interest.

The exact same mixing procedure has been applied to the PYTHIA 8 and HERWIG++ samples using the `CUETP8S1` and `CUETHS1` tunes, respectively.

A more workable form of the DPS cross section has been obtained. However, both $\sigma_{\mathrm{A,B}}^{DPS}$ and σ_{eff} are unknown. The DPS cross section still needs to be extracted from the data, before any prediction of σ_{eff} can be made.

8.1.2 Template Fit

The fraction of DPS events f_{DPS} can be extracted by performing a template fit to the fully corrected ΔS distribution. The MC models that do not contain an explicit DPS contribution are taken as the background or SPS templates. The signal template is constructed from the earlier mentioned pure DPS data sample. The distribution of the ΔS observable will be determined from the pure DPS data sample and will be corrected to generator level. From hence on forth we will make the distinction between ΔS and $\Delta \mathrm{S}_{\mathrm{DPS}}$, where the subscript DPS indicates that the ΔS distribution has been extracted from the pure DPS data sample.

The signal and background template are normalized to the cross section of the ΔS distribution obtained from the four jet data sample. The optimal value of the DPS fraction f_{DPS} is determined by performing a template fit based on a maximum likelihood technique using Poisson statistics [176]:

$$\sigma^{\text{Data}}(\Delta\text{S}) = f_{\text{DPS}} \cdot \sigma^{\text{Data}}_{\text{DPS}}(\Delta\text{S}_{\text{DPS}}) + (1 - f_{\text{DPS}}) \cdot \sigma^{\text{MC}}_{\text{SPS}}(\Delta\text{S}) \tag{8.6}$$

With the DPS fraction f_{DPS} known, the cross section $\sigma^{\text{DPS}}_{\text{A,B}}$ is determined from the integral of the integral of the differential cross section as a function ofthe ΔS observable, scaled with f_{DPS} or

$$\sigma^{\text{DPS}}_{\text{A,B}} = f_{\text{DPS}} \cdot \int \sigma^{\text{Data}}(\Delta\text{S}) d(\Delta\text{S}) \tag{8.7}$$

The result in combination with the rewritten DPS pocket formula in eq. (8.5) allows for the extraction of the effective cross sections σ_{eff}.

8.2 Shape Comparison Method

The DPS cross section $\sigma^{\text{DPS}}_{\text{A,B}}$ can be estimated through comparison of the normalized distributions of the variable ΔS from data and SPS MC models. The distributions are both normalized to the last bin at π where the least DPS-sensitivity is expected. The difference in shape between the SPS MC models and the data, respectively, can be seen as the missing DPS fraction needed to improve the models. However, in this approach one implicitly assumes that the DPS contribution in the bin at π is negligible. In the generator level study performed in Sec. 3.8, a DPS contribution of $\sim 6\%$ has been found in the bin at π, compared to a 35% DPS contribution in the first bin of the ΔS distribution. The ratio of the cross sections of the MC models over the cross section of the fully corrected data for the normalized histograms can be seen as the fraction of SPS events. The fraction of DPS events f_{DPS} is written in terms of these normalized distributions as

$$f_{\text{DPS}} = 1 - \frac{\sigma^{\text{MC}}_{\text{Norm}}(\Delta\text{S})}{\sigma^{\text{Data}}_{\text{Norm}}(\Delta\text{S})} \tag{8.8}$$

The subscript “Norm” indicates the usage of the normalized ΔS distributions. The DPS cross section $\sigma^{\text{DPS}}_{\text{A,B}}$ is estimated as the product of the fraction of DPS events with the integral over the differential cross section as a function ofthe ΔS observable or

$$\sigma^{\text{DPS}}_{\text{A,B}} = f_{\text{DPS}} \cdot \int \sigma^{\text{Data}}(\Delta\text{S}) d(\Delta\text{S}) \tag{8.9}$$

As with the template method, the combination of the result with the rewritten DPS pocket formula in eq. (8.5) allows for the extraction of the effective cross section σ_{eff}.

The shape comparison method is more straightforward and perhaps more naive compared to the template method or the tuning method. The results are expected to be more sensitive to small bin-to-bin differences in the distributions, not only in the DPS-sensitive area, but in the SPS-dominated region of the observable as well.

Chapter 9

Study of Detector Effects

In this chapter, the remaining detector effects are examined. The first effect is the trigger inefficiency, which needs to be corrected for. Afterwards, the resolution of all the variables is measured in order to determine an optimal binning for the final distributions. Lastly, the variables are compared with the PYTHIA 8 and HERWIG++ MC models at detector level, using their optimized binning.

9.1 Trigger Efficiency Correction

Whenever a jet is properly reconstructed in the whole trigger system, while meeting the p_{T} thresholds for a given jet trigger, the trigger is fired. If an event is deemed of proper quality by the DQM, the event is written to tape. However, it can be that an event of interest is not properly detected or reconstructed, leading to the trigger not being fired and the event of interest being discarded. This is referred to as the *trigger inefficiency*. The regions in the phase space where the trigger is not fully efficient need to be corrected for the observed inefficiencies.

The trigger efficiency ϵ_{trig} of all the jet triggers has been studied as a function of the jet p_{T}. The efficiency of all the triggers has been calculated by the so-called reference trigger method. Given a certain trigger of which one wants to determine the efficiency, one compares the firing of the trigger with that of a trigger with a lower p_{T} threshold or a Minimum Bias (MB) trigger. Where a MB trigger is used to select MB events, defined as single diffractive events, double diffractive events and non-diffractive events. The efficiency can be written as

$$\epsilon_{\mathrm{trig}} = \frac{N\left(\mathrm{trig} + \mathrm{trig}_{\mathrm{ref}} \text{ fired}\right)}{N\left(\mathrm{trig}_{\mathrm{ref}} \text{ fired}\right)} \tag{9.1}$$

Where $N(\ldots)$ is the amount of events selected by the given trigger(s). Important to note is that the $N(\mathrm{trig} + \mathrm{trig}_{\mathrm{ref}} \text{ fired})$ has been corrected for the prescale of the considered trigger. Whereas a correction for the prescale of the reference trigger is not necessary since the prescale cancels in the numerator and the denominator.

The `HLT_L1MinimumBiasHF_OR_partX_v1` triggers have been used as the reference trigger in combination with the `L1MinimumBiasX` data samples, where $X \in \{0, \ldots, 9\}$.

Each efficiency curve has been fitted with the error function given below.

$$\epsilon_{\text{trig}} = \frac{1}{2}\left(1 + \text{Erf}\left(\frac{p_{\text{T}} - p_0}{\sqrt{2} \cdot p_1}\right)\right) \tag{9.2}$$

Where the parameters p_0 and p_1 are determined through the fitting procedure for each trigger individually. A trigger is considered to be fully effective when the efficiency becomes larger than 0.99. This point has been determined for all the triggers and is shown in Tab. 9.1 along with the p_{T} threshold of all triggers. The resulting efficiency curves for all the triggers are shown in Fig. 9.1 and Fig. 9.2.

It is preferred to use a trigger where it is fully efficient. However, the analysis of the JEC uses a $p_{\text{T,Ave}}$ as low as 20 Gev and the inclusive four jet analysis puts a lower boundary on the leading jet p_{T} of 35 GeV. The triggers used in the low p_{T} regions are not always fully efficient, therefore, a trigger correction needs to be applied in the low p_{T} regions of the phase space, effectively altering the weight of the events. The regions of the phase space where the different triggers have been used can be found in Sec. 6.2.2 and Sec. 7 for the determination of the JEC and for the inclusive four jet analysis, respectively.

Table 9.1: List of all the triggers with their respective p_{T} thresholds and the values where they become fully efficient. The threshold for the `DiPFJet` triggers is the p_{T} threshold for the second leading jet instead of the leading jet, while the threshold for the `DiPFJetAve` triggers is the $p_{\text{T,Ave}}$ of the two leading jets.

Trigger	p_{T} threshold (GeV)	Full efficiency threshold (GeV)
Triggers used for the JEC		
`HLT_PFJet15_NoCaloMatched_v3`	15	30
`HLT_PFJet25_NoCaloMatched_v1`	25	35
`HLT_DiPFJet15_NoCaloMatched_v1`	15	26
`HLT_DiPFJet25_NoCaloMatched_v1`	25	37
`HLT_DiPFAveJet15_HFJEC_v1`	15	24
`HLT_DiPFAveJet25_HFJEC_v1`	25	34
`HLT_DiPFAveJet35_HFJEC_v1`	35	52
Triggers used in the 4-jet analysis		
`HLT_AK4PFJet30_v3`	30	43
`HLT_AK4PFJet50_v3`	50	64
`HLT_AK4PFJet80_v3`	80	97
`HLT_AK4PFJet100_v3`	100	125

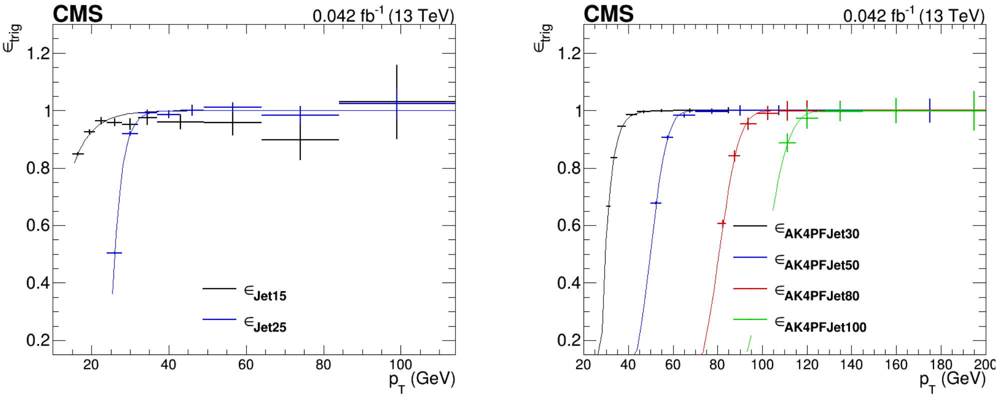


Figure 9.1: The trigger efficiencies as a function of the p_T of the leading jet for the `PFJet` (left) and `AK4PFJet` (right) triggers.

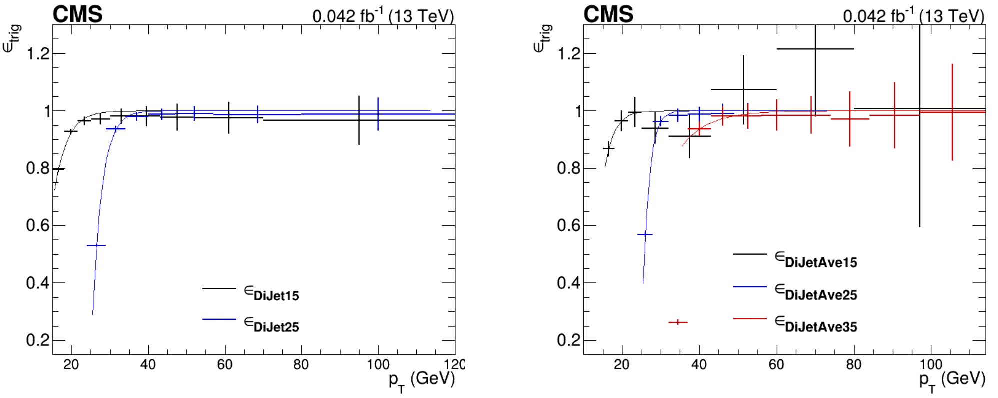


Figure 9.2: The trigger efficiencies for the different `DiPFJet` (left) and `DiPFJetAve` (right) jet triggers. The efficiency is as a function of the p_T of the sub-leading jet for the `DiPFJet` triggers, while the efficiency curves for the `DiPFJetAve` triggers (right) are as a function of the $p_{T,\mathrm{Ave}}$.

9.2 Detector Resolution

The CMS detector, just as any measuring device, has a finite resolution. A detector is never able to determine or measure the exact value of an angle, energy deposit, … Measurements performed by the CMS detector are affected by such effects as calibration uncertainties, the non-linearity of the detector response, … A particle with a true energy E can be measured with a different energy E^* by the detector. In general, the measured energies E^* will form a peaked distribution centered around the true energy E. The half-width of the distribution is referred to as the resolution, which is a measure of the preciseness and accurateness of the detector.

The resolution of a variable plays an important role in the determination of the binning of the histogram for said variable. Due to a non-zero resolution, measured values can migrate to a different bin compared to the bin of their true value. Migrations should be avoided as they can complicate the unfolding procedure, i.e. correcting for the remainder of detector effects. The procedure is detailed in Sec. 10. As a rule of thumb to counter bin-to-bin migrations, the bin width is taken to be at least 2 times the resolution of the bin. The smearing into neighboring bins will be smaller than 5% in the case of a Gaussian distributions.

The PYTHIA 8 sample has been used for the determination of the resolution of all observables. Apart from the standard event selection, additional matching criteria are put in place. Jets at detector and generator level are matched and the resolution of the variables is calculated only for events with at least 4 matched jets. The matching procedure varies for different observables as described below.

- **p_{T}-based observables:** The jets are matched in ϕ and η where

  $$\sqrt{|\phi^{\mathrm{gen}} - \phi^{\mathrm{det}}|^2 + |\eta^{\mathrm{gen}} - \eta^{\mathrm{det}}|^2} \leq \sqrt{R_\phi^2 + R_\eta^2} \tag{9.3}$$

  The maximum separations in ϕ and η are $R_\phi = 0.1$ and $R_\eta = 0.1$. The p_{T}-based observables are the p_{T} spectra of the four leading jets, $\Delta p_{\mathrm{T,soft}}$ and $\Delta\mathrm{S}$.

- **ϕ-based observables:** The jets are matched in p_{T} and η where

  $$\left|\frac{p_{\mathrm{T}}^{\mathrm{gen}} - p_{\mathrm{T}}^{\mathrm{det}}}{p_{\mathrm{T}}^{\mathrm{gen}}}\right| \leq R_{p_{\mathrm{T}}} \quad \text{and} \quad \left|\eta^{\mathrm{gen}} - \eta^{\mathrm{det}}\right| \leq R_\eta \tag{9.4}$$

  The maximum separations in p_{T} and η are $R_{p_{\mathrm{T}}} = 0.3$ and $R_\eta = 0.1$. The ϕ-based observables are $\Delta\phi_{\mathrm{soft}}$, $\Delta\phi_{\mathrm{3j}}^{\mathrm{min}}$, ϕ_{ij}.

- **η-based observables:** The jets are matched in p_{T} and ϕ where

  $$\left|\frac{p_{\mathrm{T}}^{\mathrm{gen}} - p_{\mathrm{T}}^{\mathrm{det}}}{p_{\mathrm{T}}^{\mathrm{gen}}}\right| \leq R_{p_{\mathrm{T}}} \quad \text{and} \quad \left|\phi^{\mathrm{gen}} - \phi^{\mathrm{det}}\right| \leq R_\phi \tag{9.5}$$

  The maximum separations in p_{T} and η are $R_{p_{\mathrm{T}}} = 0.3$ and $R_\eta = 0.1$. The η-based observables are all the pseudorapidity spectra, i.e. the four leading jets of the inclusive four jet production and the leading jets of the inclusive single jet processes, and $\Delta\mathrm{Y}$.

The resolution of the p_{T} of the jets is relative since the p_{T} of the jets ranges from 20 GeV up to around 450 GeV. The radii R_{ϕ}, R_{η} and $R_{p_{\mathrm{T}}}$ have been determined individually by examining the difference of the azimuthal angle, of the pseudorapidity and of the transverse momentum, respectively, between all detector and generator level jets in each event, resulting in three peaked distributions centered around zero. The radii correspond to where the distributions started to rise in order to assure that jets correlated on detector and generator level were being compared to one another. The values of $R_{p_{\mathrm{T}}}$, R_{η} and R_{ϕ} have been varied up and down to justify their value. The result was a decrease in statistics while the resolution remained the same, if smaller values were used. Whereas the resolution decreased if the matching conditions were loosened since more and more decorrelated jets would be allowed in the calculation of the resolution. By implementing the matching procedure, the resolution of the observable in question is determined independent of the matching procedure. In other words, if the matching for p_{T}-based observables would entail a matching in p_{T}, then correlations between the p_{T} of the jets and the radius $R_{p_{\mathrm{T}}}$ can occur. The resolution would not be independent of the matching procedure.

The resolution together with an appropriate detector level binning have been determined for all variables. The generator level binning will always be coarser compared to the detector level binning. The algorithm of the unfolding procedure requires that the number of detector level bins is strictly larger than the number of generator bins, see Sec. 10. If the rule of thumb is met at detector level, it will automatically be met at generator level. Fig. 9.3, Fig. 9.4 and Fig. 9.5 depict the resolution for all observables.

The resolution of the p_{T} spectra of the four leading jets varies between 7-15% and increases when the p_{T} of the jet increases. For the pseudorapidity spectra of all the jets, the maximum observed resolution is 0.06. A small increase is seen when comparing the resolution of the first to the last leading jet, which is related to the differences in the p_{T} spectra. More low p_{T} jets enter the p_{T} spectrum of the fourth leading jet compared to the last leading jet, resulting in a slightly worse resolution. As the differences in resolution between the jets are small, the same pseudorapidity binning will be used for all the pseudorapidity spectra. The inclusive single jet cross sections σ_{A} and σ_{B} will be measured from their pseudorapidity spectra η_{A} and η_{B} and will use the same binning as all other pseudorapidity spectra.

All variables that are not matched in p_{T}, i.e. $\Delta\phi_{\mathrm{soft}}$, $\Delta\phi_{3\mathrm{j}}^{\mathrm{min}}$, ϕ_{ij} and $\Delta\mathrm{Y}$, exhibit a resolution between 0.04 and 0.06. The resolution of the variable $\Delta p_{\mathrm{T,soft}}$ increases as $\Delta p_{\mathrm{T,soft}}$ increases, which can be understood from its definition in eq. (3.21). When the two softest jets lie in a back-to-back topology the numerator will be small. The value of $\Delta p_{\mathrm{T,soft}}$ will be more susceptible to variations between the p_{T} of the two jets on detector and generator level, resulting in a broader distribution.

The resolution of $\Delta\mathrm{S}$ is much larger than for all other variables even with the stricter cuts on the p_{T} of the jets. The $\Delta\mathrm{S}$ variable is the only one for which the rule of thumb is not met in all bins. The reason why this detector level binning is chosen and not further altered in order to meet the rule of thumb, is two-fold. First, the p_{T} of all four jets enters the definition of $\Delta\mathrm{S}$ in eq. (3.26), yielding a high sensitivity to any small deviation between the p_{T} of the jets on detector and generator level. Increasing the cuts on the p_{T} of the jets further, will result in a better resolution but fewer statistics, affecting the stability of the unfolding procedure. Second, reducing the number of bins on detector level will result in a loss of sensitivity. The final binning for the variable $\Delta\mathrm{S}$ has been determined by balancing the increase in resolution from raising the cuts with the amount of statistics necessary for a stable unfolding procedure. The $\Delta\mathrm{S}_{\mathrm{DPS}}$ observable, extracted from the pure DPS sample, will use the same binning as the $\Delta\mathrm{S}$ observable since both observables will be used in an effort to extract the DPS cross section and the effective cross section.

The final binning of all observables at detector level is detailed below.

- $p_{\mathrm{T,1st}} = [35, 45, 60, 75, 95, 120, 150, 200, 250, 300, 375, 450]$
- $p_{\mathrm{T,2nd}} = [30, 45, 60, 75, 95, 120, 150, 200, 250, 300, 375]$
- $p_{\mathrm{T,3rd}} = [25, 32.5, 45, 60, 80, 105, 150, 200, 250, 350]$
- $p_{\mathrm{T,4th}} = [20, 27.5, 35, 45, 60, 80, 105, 150, 200, 280]$
- $\eta_i, \eta_{\mathrm{A}}, \eta_{\mathrm{B}} = [-4.716, -3.839, -3.489, -2.964, -2.5, -2.172, -1.93, -1.74, -1.566, -1.305, -1.131, -0.957, -0.609, -0.261, 0.261, 0.609, 0.957, 1.131, 1.305, 1.566, 1.74, 1.93, 2.172, 2.5, 2.964, 3.489, 3.839, 4.716]$
- $\Delta\phi_{\mathrm{soft}}, \Delta\phi_{\mathrm{3j}}^{\mathrm{min}}, \phi_{\mathrm{ij}} = [0, 0.020, 0.039, 0.059, 0.079, 0.098, 0.118, 0.137, 0.157, 0.177, 0.196, 0.216, 0.236, 0.255, 0.275, 0.295, \pi]$
- $\Delta\mathrm{Y} = [0, 0.4, 0.8, 1.15, 1.5, 2.1, 2.6, 3.1, 3.6, 4.1, 4.6, 5.1, 5.7, 6.3, 6.9, 7.8, 9.4]$
- $\Delta p_{\mathrm{T,soft}} = [0, 0.175, 0.35, 0.475, 0.6, 0.675, 0.75, 0.825, 0.9, 0.95, 1]$
- $\Delta\mathrm{S}, \Delta\mathrm{S}_{\mathrm{DPS}} = [0, 0.6, 1.2, 1.7, 2.1, 2.4, 2.65, 2.85, 3.00, \pi]$

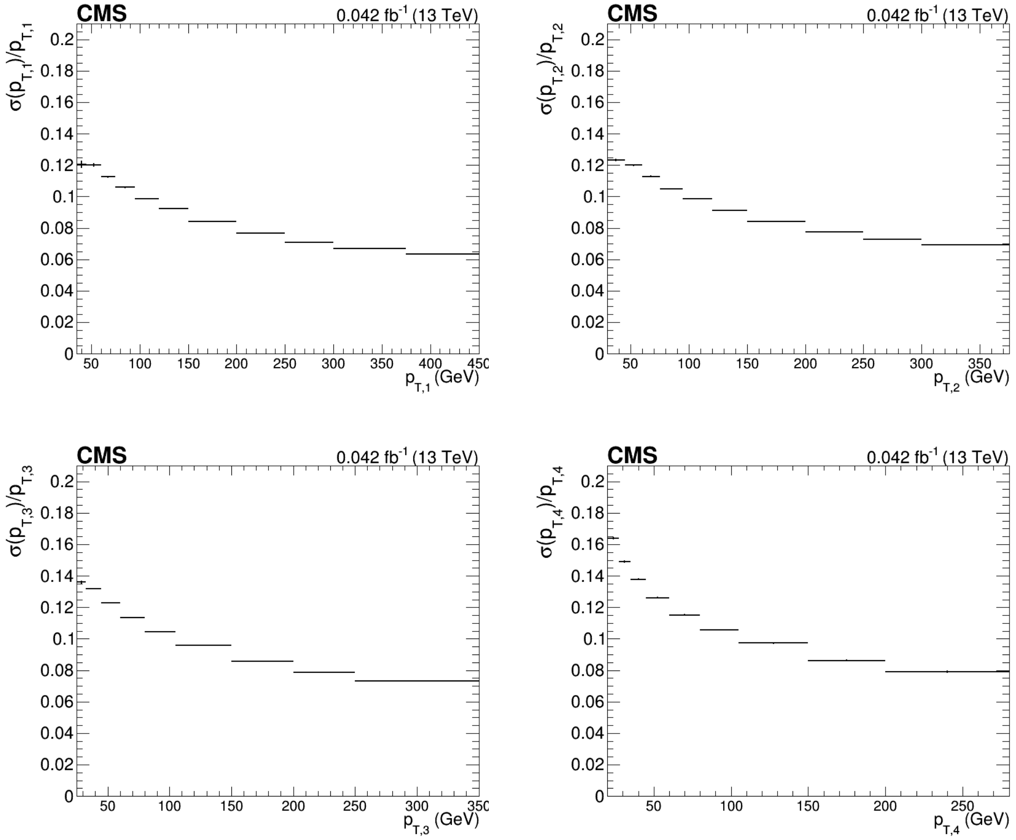


Figure 9.3: The relative resolution for the p_{T} spectra of the leading to the last leading jet are shown from the top left to the bottom right plot, respectively.

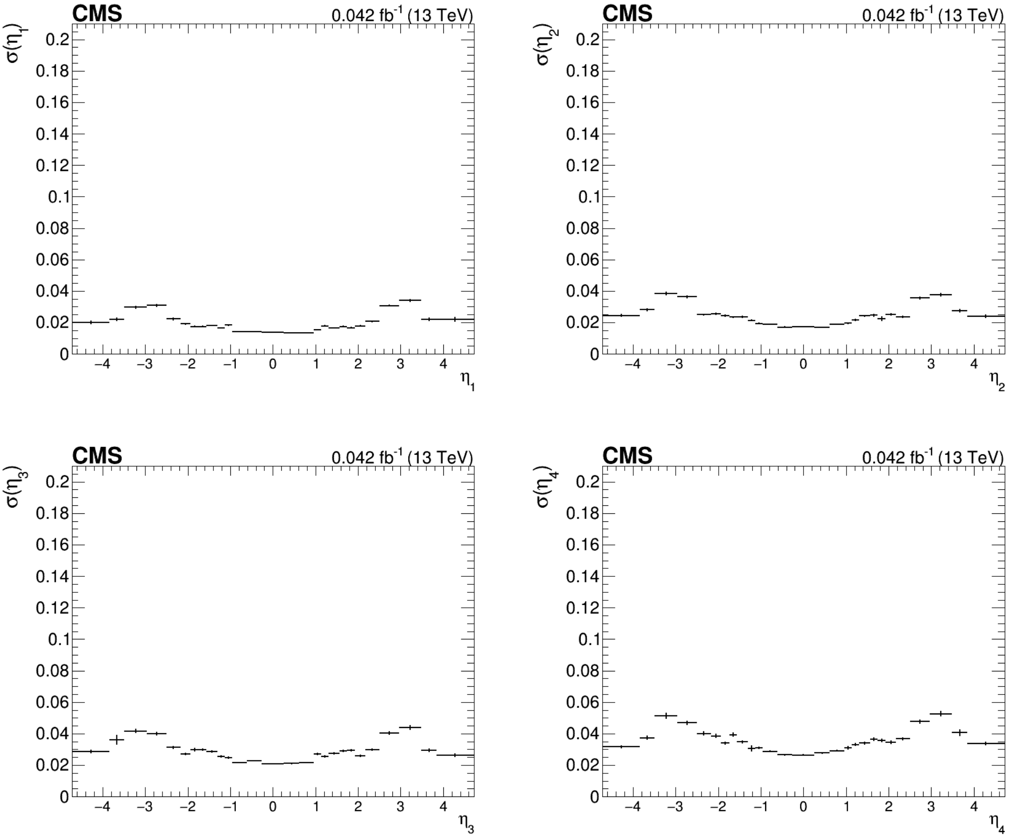


Figure 9.4: The resolution for the pseudorapidity spectra of the leading to the last leading jet are shown from the top left to the bottom right plot, respectively.

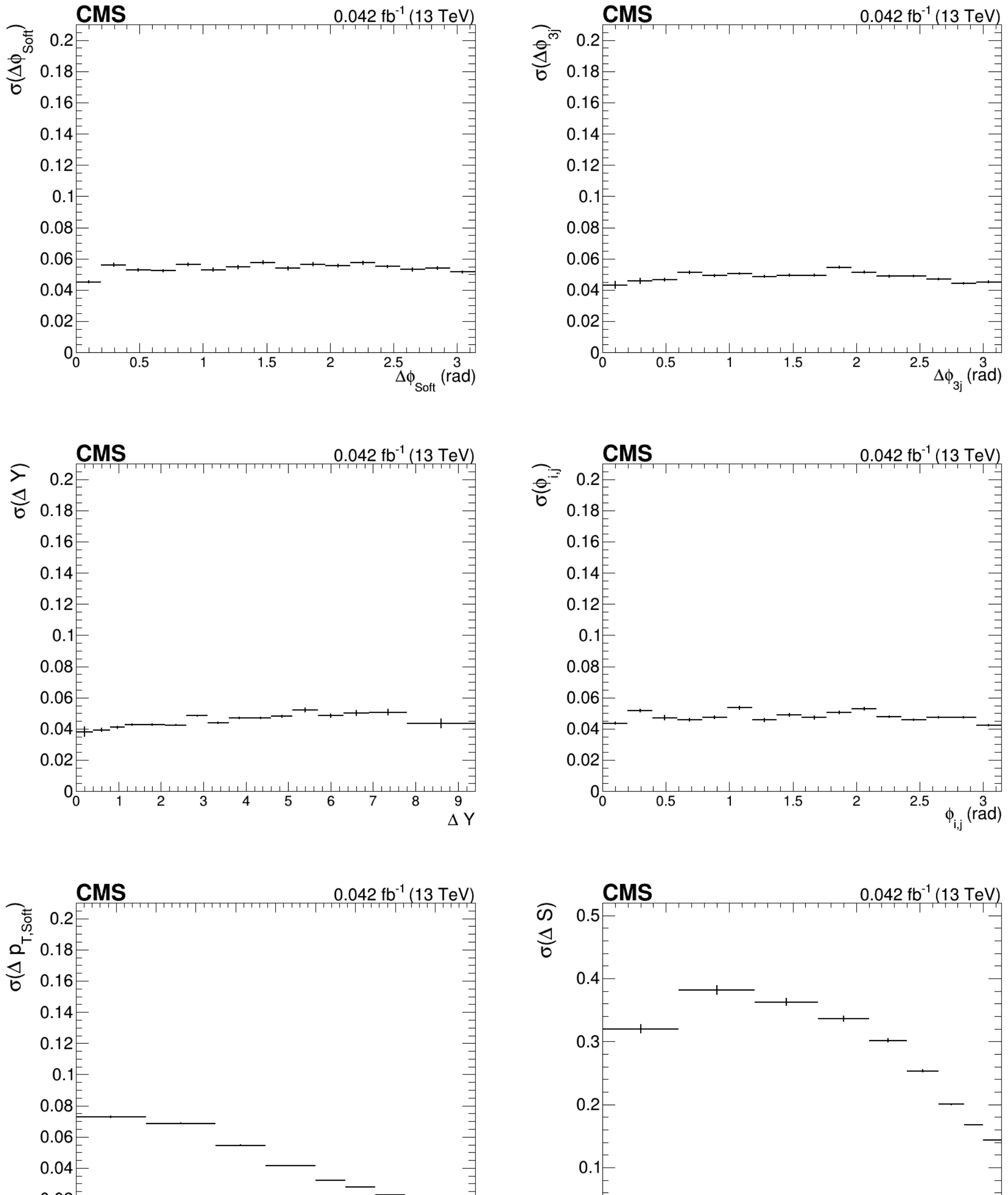


Figure 9.5: The resolution for the DPS-sensitive observables. The resolution of the ΔS variable (bottom right) has been determined in *region II*, compared to *region I* for all other observables.

9.3 Detector Level Study

As an appropriate detector level binning has been determined, all variables have been plotted at detector level together with the PYTHIA 8 and HERWIG++ samples using the `CUETP8M1` and `CUETHS1` tunes, respectively, to cross check the reliability of the detector simulation. The data has been corrected for the vertex selection criterion, in other words: the luminosity is corrected for the fraction of events with exactly one proper vertex relative to the total amount of events. The weight of events is further multiplied with an additional factor due to possible trigger inefficiencies. Tab. 9.2 lists the number of selected events after the application of the online and the offline selection criteria in succession. The second to last column is the total amount events used in this analysis for all variables except $\Delta\mathrm{S}$, which is given in the last column.

Table 9.2: The number of events surviving the online and offline selection criteria applied in succession from left to right. Once the soft (hard) kinematic cuts have been applied, the phase space region corresponds to *region I* (*region II*).

Sample	Trigger	Total number of events	Trigger selection	Vertex selection	Soft kinematic cuts (*region I*)	Hard kinematic cuts (*region II*)
`HINPFJets`	`AK4PFJets30`	5,905,759	4,433,002	3,173,906	52,731	5,552
	`AK4PFJets50`	5,905,759	170,818	121,940	20,526	6,494
	`AK4PFJets80`	5,905,759	29,253	20,579	6,580	3,094
	`AK4PFJets100`	5,905,759	15,917	11,114	5,009	3,068
PYTHIA 8	-	9,700,000	-	9,421,695	3,567,120	3,303,214
HERWIG++	-	9,993,000	-	9,730,687	3,558,557	3,255,406

The p_T and η spectra are shown in Fig. 9.6 and Fig. 9.7, respectively, while the observables are depicted in Fig. 9.8 and Fig. 9.9.

Apart from an overall difference in cross section, the p_T spectra of the data show a different behavior at low p_T compared to the MC samples. The MC samples overshoot the data in the low p_T region of phase space. At high p_T, the ratio of the MC samples over the data becomes constant. In this region of the phase space the only difference between the models is the overall cross section. For the pseudorapidity spectra of the four leading jets, a difference in shape is observed in the forward and backward regions. The discrepancy is linked to the overshoot of low p_T jets. Forward and backward jets carry on average a lower transverse momentum compared to central jets.

The shape of the variables $\Delta\phi_\mathrm{soft}$, $\Delta p_\mathrm{T,soft}$ and $\Delta\mathrm{S}$ is relatively well described by the PYTHIA 8 and HERWIG++ MC samples as the ratios remain rather homogeneous. The shape of the observable $\Delta\phi_\mathrm{3j}^\mathrm{min}$ lies somewhere in between the two MC models as the ratios relative to both models cross each other around $\Delta\phi_\mathrm{3j}^\mathrm{min}$ equal to 2 rad. A large deviation in shape is observed between the data and the models for the variable $\Delta\mathrm{Y}$. Due to an excess of forward and backward jets in the MC models, the average maximum separation in pseudorapidity is larger as in the data, resulting in the large overshoot at high values of $\Delta\mathrm{Y}$. The observable ϕ_ij is relatively well described by the HERWIG++ model, while the PYTHIA 8 sample greatly overshoots the shape of the distribution at low values of ϕ_ij.

The pseudorapidity spectra of the inclusive single jet cross sections of the MC samples overshoot the data. The discrepancy is larger for the process B compared to process A since jets with a lower p_T are included in σ_B compared σ_A. The detector level $\Delta\mathrm{S}_\mathrm{DPS}$ distribution, shown in the bottom plot of Fig. 9.9, displays a much more decorrelated picture as the ΔS observable obtained from the data. All the ΔS_{DPS} distributions have been scaled to the uncorrected detector level cross section of the observable ΔS. The cross sections of the ΔS_{DPS} observable obtained from the pure DPS data and MC samples are not known.

Both MC samples use a $2 \rightarrow 2$ ME, therefore at least two of the selected jets will come from the parton shower. The main difference between the PYTHIA 8 and HERWIG++ models is the type of parton shower. Whereas PYTHIA 8 has implemented a p_T-ordered shower, HERWIG++ uses an angular ordered shower. The different treatment of partons in the parton showers could lie at the root of the discrepancies observed in the spectra of the $\Delta\phi_{3j}^{\min}$ and ϕ_{ij} observables for example.

The PYTHIA 8 and HERWIG++ models fail to describe the complete phase space at detector level, however, they will still be used to fully correct the data to generator level. Additional steps will be undertaken in Sec. 10 to assure that the migrations relative to the data are properly described by the MC models.

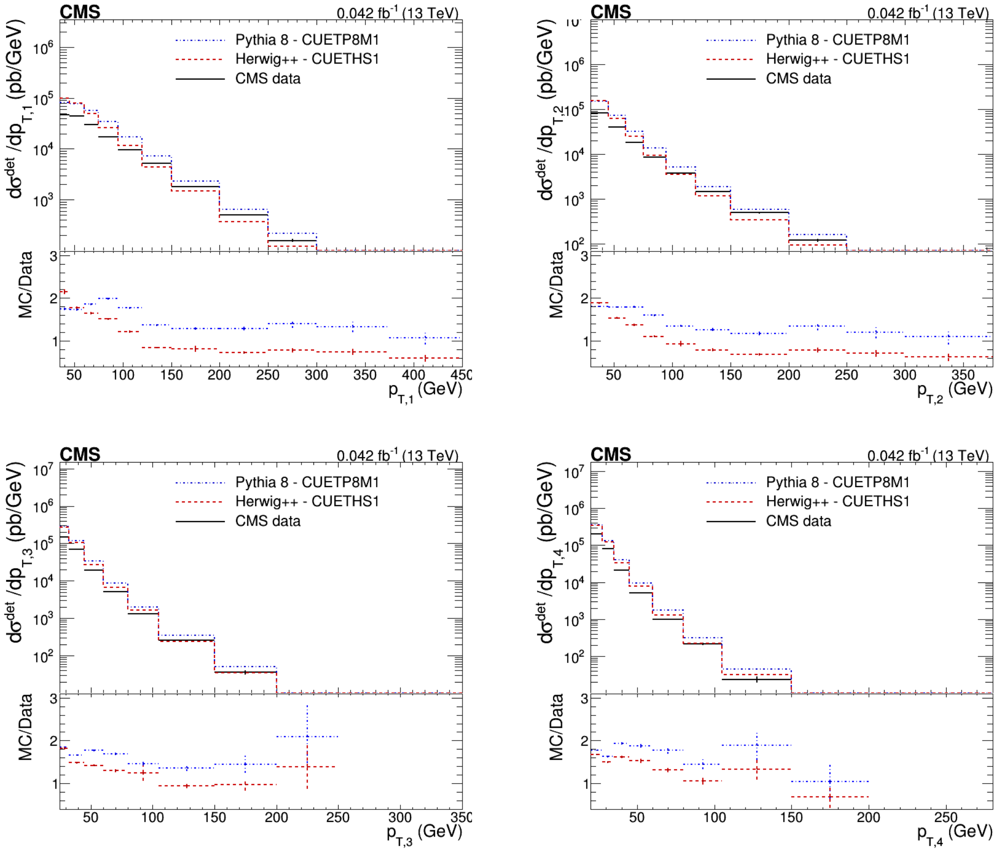


Figure 9.6: The detector level data along with the PYTHIA 8 and HERWIG++ MC predictions of the p_T spectra for the leading to the last leading jet are shown from the top left to the bottom right plot, respectively.

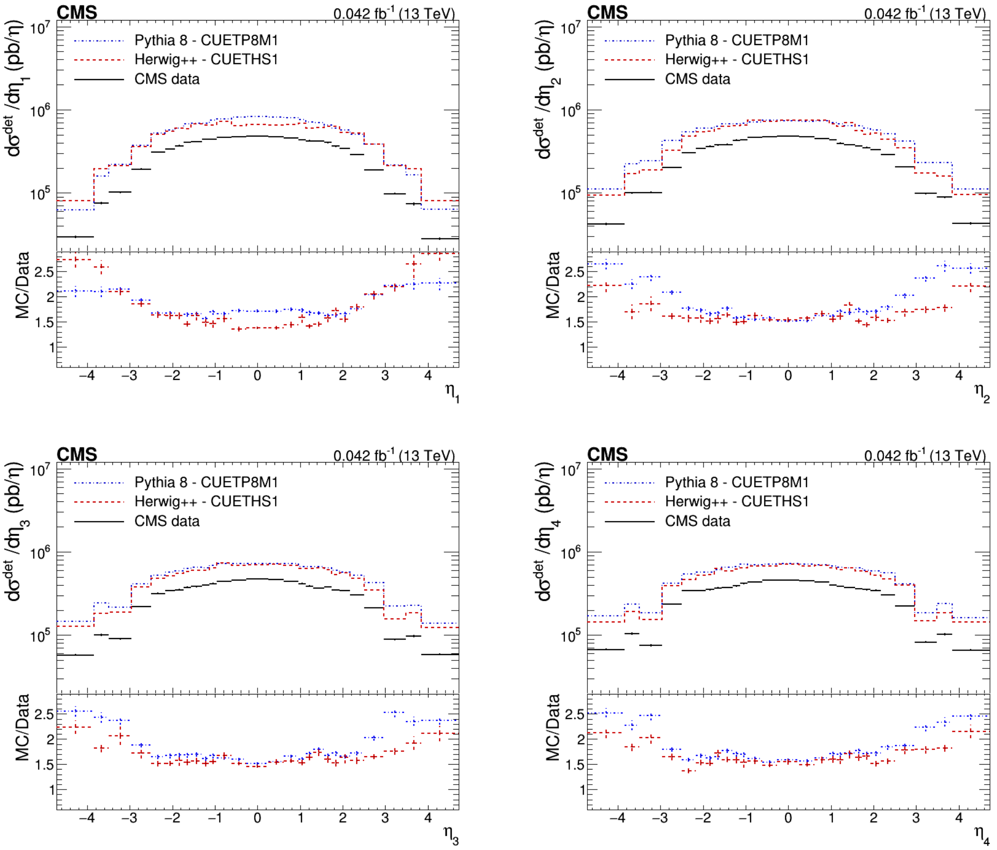


Figure 9.7: The detector level data along with the PYTHIA 8 and HERWIG++ MC predictions of the η spectra for the leading to the last leading jet are shown from the top left to the bottom right plot, respectively.

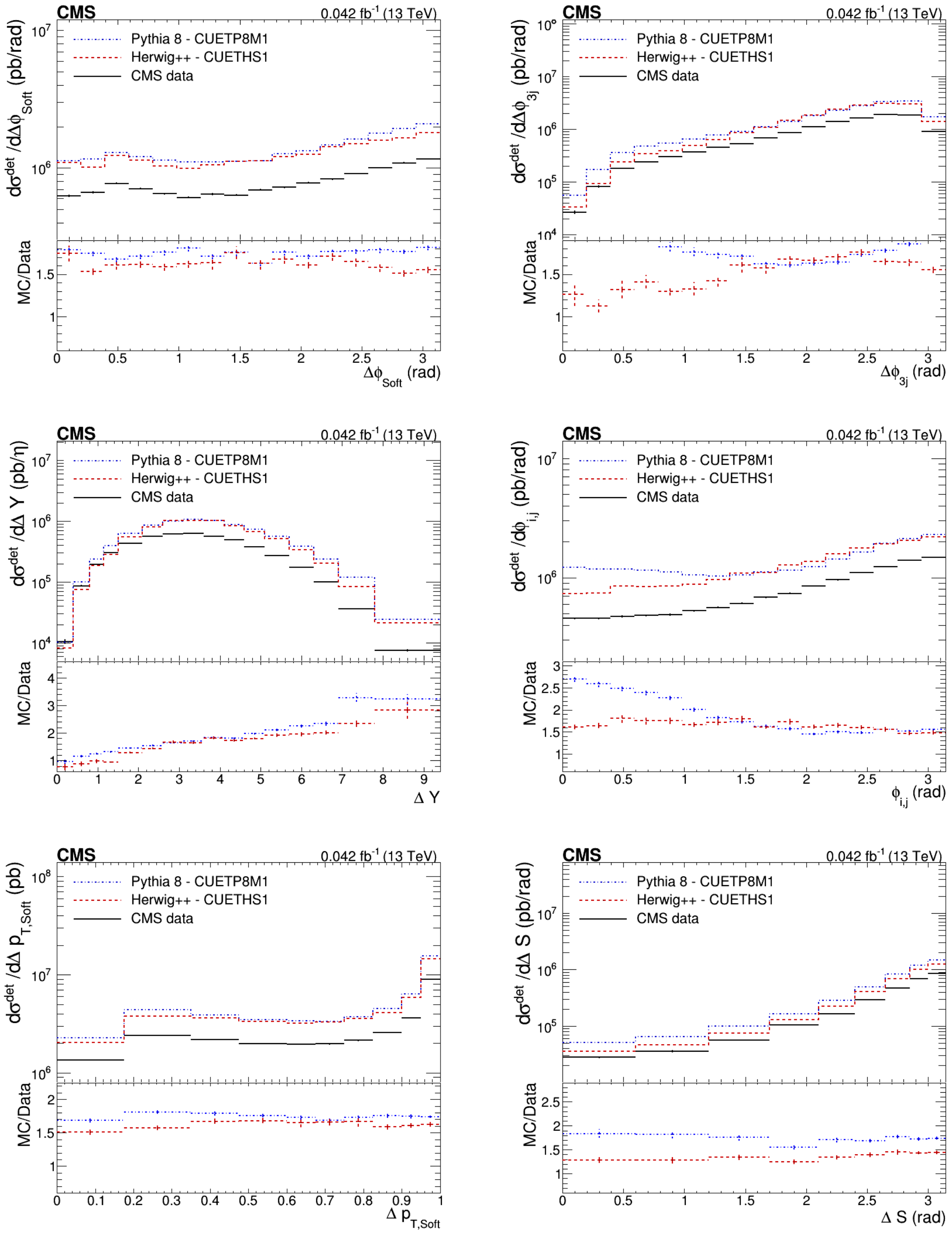


Figure 9.8: The detector level data along with the PYTHIA 8 and HERWIG++ MC predictions for the DPS-sensitive observables. The ΔS observable (bottom right) has been determined in *region II*, compared to *region I* for all other observables.

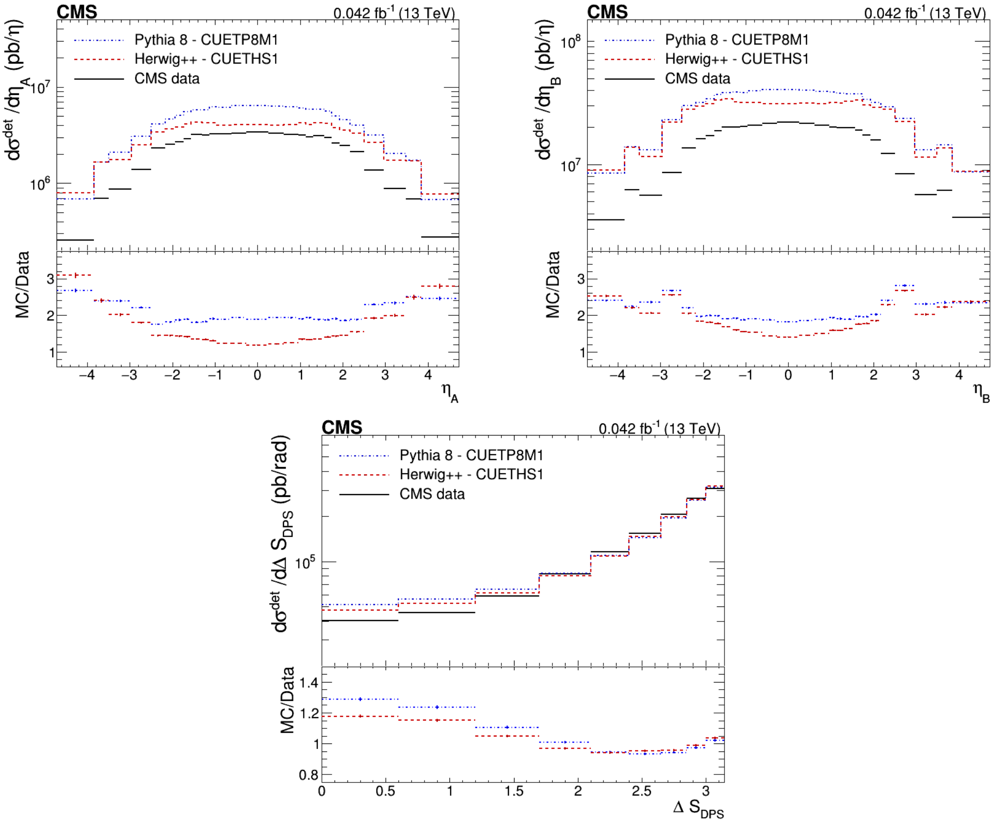


Figure 9.9: The detector level data along with the PYTHIA 8 and HERWIG++ predictions for the pseudorapidity spectra of the inclusive single jet processes A and B are shown on the top left and top right plots, respectively. The detector level ΔS_{DPS} distribution for the mixed data and mixed MC samples is depicted on the bottom plot. The ΔS_{DPS} distribution has been determined in *region II* and has been scaled to the detector level cross section of the ΔS observable.

Chapter 10

Unfolding the Data

A measurement is subjected to inevitable detector effects, causing differences on generator and detector level, characterized by the migrations. Different unfolding procedures can be used to correct for these migrations. It is opted to use an unfolding method based on the Singular Value Decomposition (SVD) of a matrix [177] along with Tikhonov regularization [178]. First, an overview of the unfolding method and all elements that come into play will be discussed. Next, the migrations are studied, after which the unfolding procedure is performed. Finally, the unfolded plots will be shown together with multiple cross checks of the whole procedure.

10.1 General Formulation of the Unfolding Problem

In general the problem of unfolding may be formulated as

$$\tilde{\mathbf{y}} = \mathbf{A}\tilde{\mathbf{x}} + \mathbf{b} \tag{10.1}$$

where

- $\tilde{\mathbf{x}}$ is the truth vector at generator level, which is approached by the optimal estimator $\mathbf{x}$.
- $\tilde{\mathbf{y}}$ is the vector of the average expected event count per bin at detector level.
- $\mathbf{A}$ is the matrix of probabilities that describes the migrations within the phase space and effectively couples the detector and generator level to one another, see Sec. 10.2.2.
- $\mathbf{b}$ is the background vector, which will be discussed in detail in the study of the into and out of phase space migrations in Sec. 10.2.3.

One wishes to obtain the best possible estimate $\mathbf{x}$ of the truth vector $\tilde{\mathbf{x}}$. The equation is solved for $\mathbf{x}$, after the average expected distribution $\tilde{\mathbf{y}}$ is replaced by the measured distribution $\mathbf{y}$. However, inverting the problem is not guaranteed to work due to instabilities in the inversion of the matrix $\mathbf{A}$. The instabilities can be understood from the SVD of a matrix [177]. The SVD is a generalization of the eigendecomposition of a square matrix and allows any real or complex $m \times n$ matrix to be written as

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}} \tag{10.2}$$

Where $\mathbf{U}$ and $\mathbf{V}$ are m and n dimensional orthogonal matrices, respectively. The matrix $\mathbf{S}$ is a $m \times n$ rectangular diagonal matrix with positive real numbers σ_i on its diagonal. The diagonal elements are the so-called singular values. It is important to note that m and n are the number of bins of the detector and generator level, respectively, as the matrix $\mathbf{A}$ relates both levels to one another.

As the inverse of an orthogonal matrix is its own transpose, the problem of the inversion of the matrix $\mathbf{A}$ is directly connected to the inverse of the determinant of the matrix $\mathbf{S}$ or $\det\mathbf{S} = \prod \sigma_i$. The problem of the inversion of the matrix $\mathbf{A}$ through the SVD can be ill-posed in two cases.

- If the number of bins on generator level n exceeds the number of bins on detector level m. In this case, an infinite number of solutions that fit the data equally well exist.
- If the ratio of the largest and smallest singular value is large, then the matrix inversion is dominated by the small singular values, yielding an unstable inversion. The stability of the inversion or how well-/ill-posed the problem is, can be quantified by the condition number defined below.

  $$\mathrm{cond}(\mathbf{A}) = \frac{\sigma_{\mathrm{max}}}{\max(0, \sigma_{\mathrm{min}})} \tag{10.3}$$

  Where σ_{max} and σ_{min} represent the largest and smallest singular values of the SVD. Singular values can become negative due to a finite precision in calculations. Therefore, the maximum between the smallest value and zero is taken in the denominator. If the condition number is small, i.e. of the order of 10, then the problem of inversion is well-posed. Whenever the condition number takes on values of 10^5, it is considered large and the problem becomes ill-posed.

In order to deal with the possibility of an ill-posed problem where $n \leq m$, the unfolding algorithm [179, 180] has implemented a least-square minimization with Tikhonov regularization [178] which is discussed in detail in Sec. 10.3. A schematic overview of the unfolding procedure is depicted in Fig. 10.1.

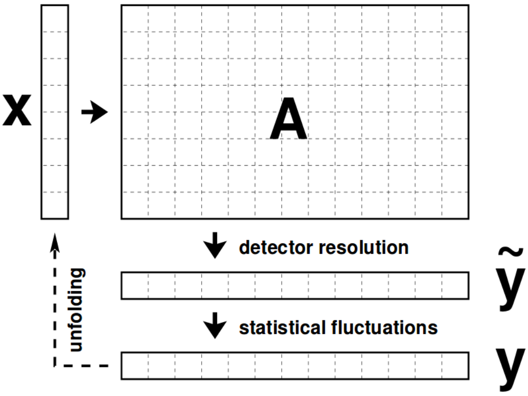


Figure 10.1: Schematic representation of the unfolding procedure and the relation between the different components, taken from [179].

10.2 Study of the Migrations

In the formal discussion of the unfolding problem, the matrix **A**, linking the generator and detector level, and the background vector **b** have been introduced. Both are described by the migrations, i.e. differences between the detector and generator level observables. The distinction between two different types of migrations is made. On the one hand, the matrix **A** describes the migrations within the phase space. On the other hand, the background vector **b** is determined by the migrations into or out of the phase space.

10.2.1 Reweighting of the MC Samples

The PYTHIA 8 and HERWIG++ MC samples will be used to correct for the remaining detector effects in the unfolding procedure. Therefore, it is important that the samples represent the data properly. If the MC samples do not describe the data accurately in shape, then the data can be overcorrected or undercorrected due to a mishandling of the migrations. In order to assure that the MC samples match the data well, a reweighting of the MC samples is performed.

The ratios between the MC samples and the data at detector level are shown in Fig. 9.6, Fig. 9.7, Fig. 9.8 and Fig. 9.9 for all observables. The reweighing function w_O for each observable O is defined as the ratio of the differential cross section as a function of the observable on detector level for the MC predictions over the one obtained from data. For each event, the reweighing function of all observables is evaluated at the observable on generator level O_{gen}. The additional weight for each observable of each event is then determined by the equation below.

$$w_O(O_{\text{gen}}) = \left(\frac{d\sigma^{\text{det}}_{\text{MC}}}{dO_{\text{det}}}\right) \Big/ \left(\frac{d\sigma^{\text{det}}_{\text{data}}}{dO_{\text{det}}}\right)(O_{\text{gen}}) \tag{10.4}$$

The full weight factor of the MC samples now becomes the generated event weight times the factor above for each observable in each event.

A comparison of the reweighted detector level MC distributions with the data is shown in Fig 10.2, Fig. 10.3, Fig. 10.4 and Fig. 10.5. The reweighted samples will be used throughout the whole unfolding procedure. The original MC samples will only be compared to the final unfolded curves when discussing the results in in Chap. 12.

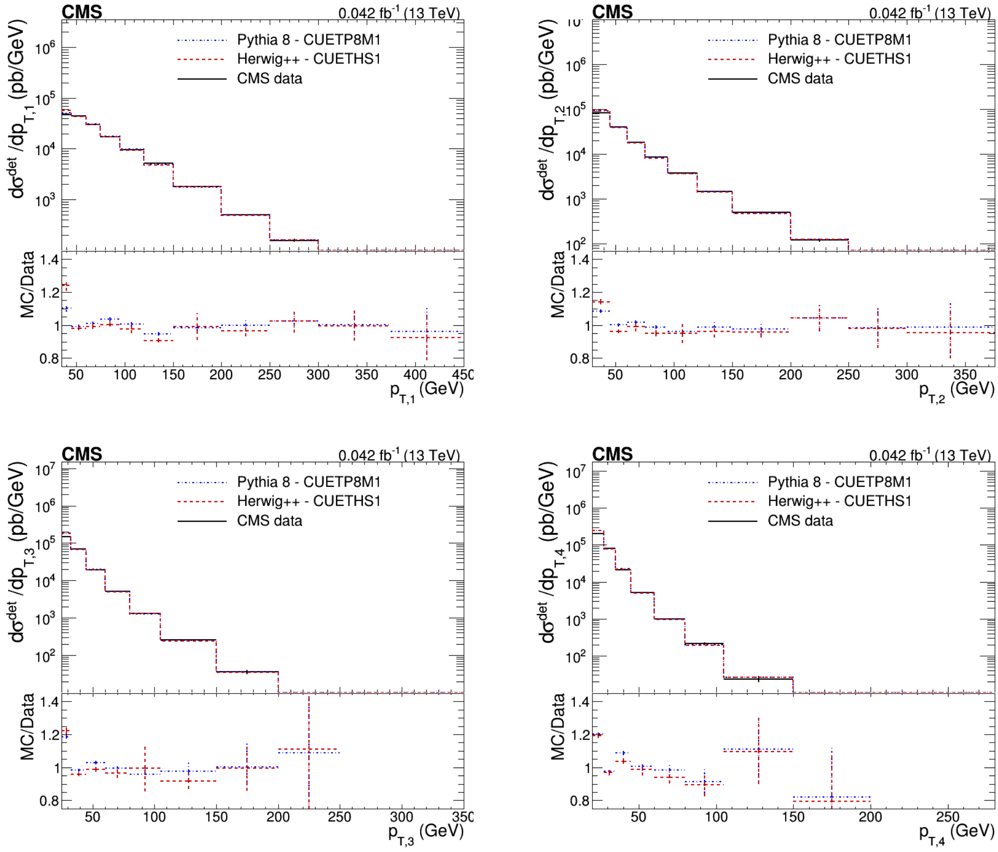


Figure 10.2: The detector level data along with the reweighted PYTHIA 8 and HERWIG++ MC predictions of the p_{T} spectra for the leading to the last leading jet are shown from the top left to the bottom right plot, respectively.

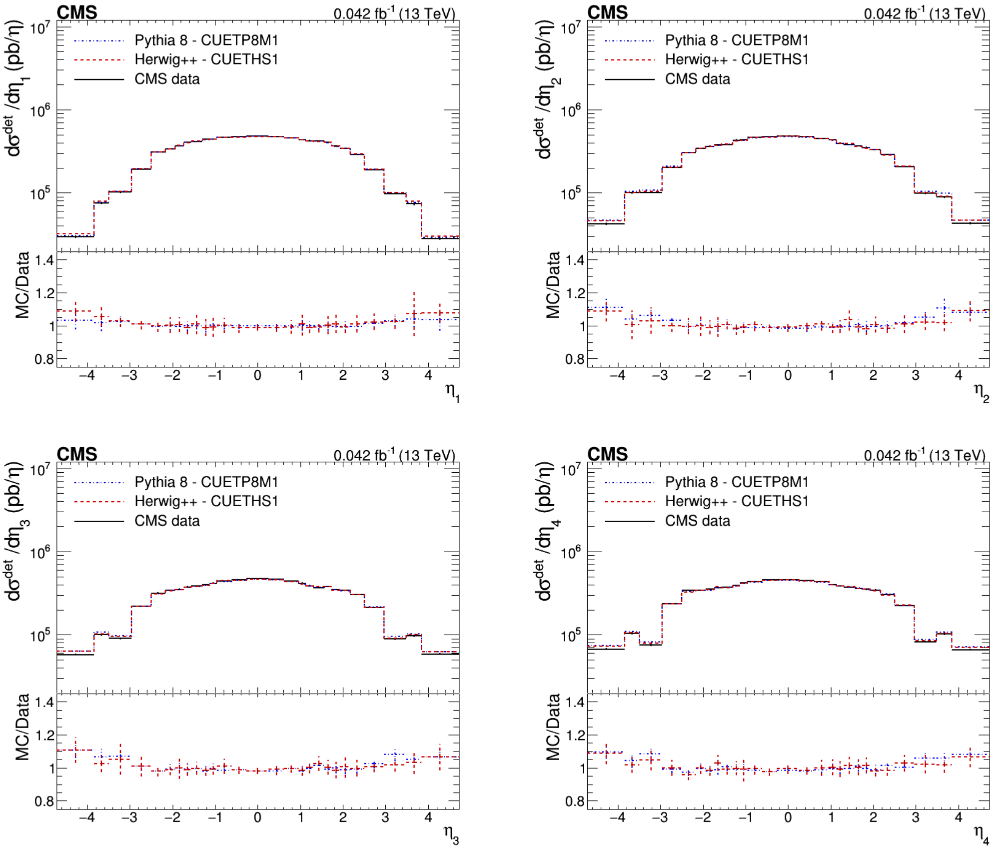


Figure 10.3: The detector level data along with the reweighted PYTHIA 8 and HERWIG++ MC predictions of the η spectra for the leading to the last leading jet are shown from the top left to the bottom right plot, respectively.

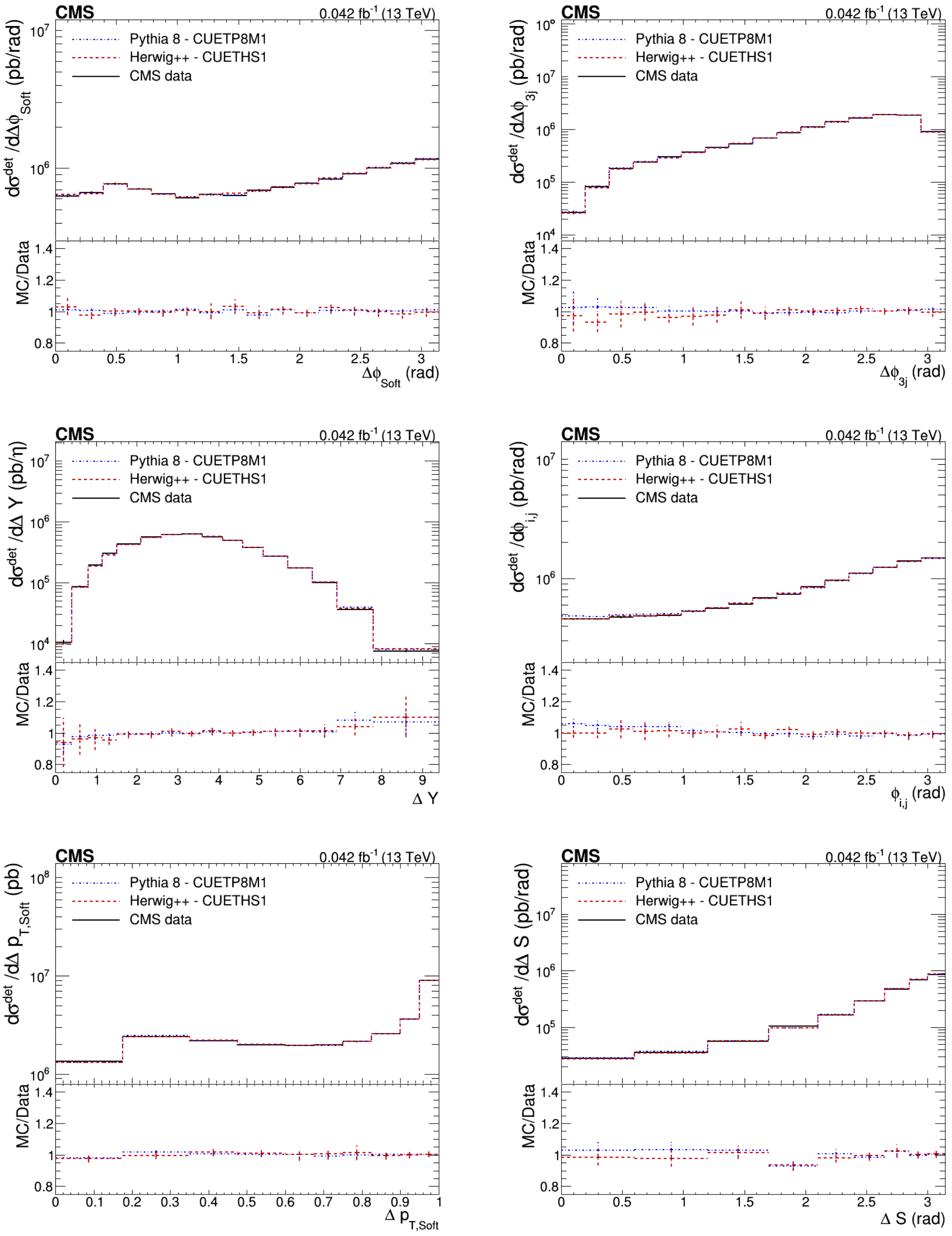


Figure 10.4: The detector level data along with the reweighted PYTHIA 8 and HERWIG++ MC predictions for the DPS-sensitive observables. The ΔS observable (bottom right) has been determined in *region II*, compared to *region I* for all other observables.

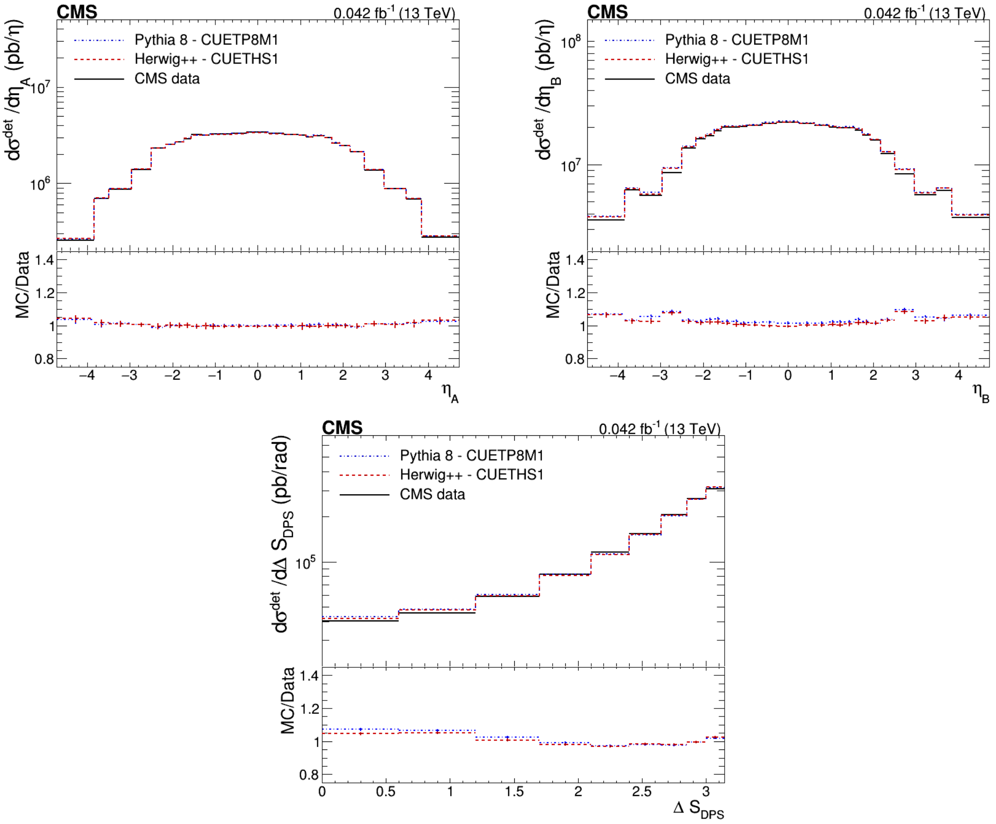


Figure 10.5: The detector level data along with the reweighted PYTHIA 8 and HERWIG++ predictions for the pseudorapidity spectra of the inclusive single jet processes A and B are shown on the top left and top right plots, respectively. The detector level $\Delta\mathrm{S}_{\mathrm{DPS}}$ distribution for the mixed data and mixed reweighted MC samples is depicted on the bottom plot. The $\Delta\mathrm{S}_{\mathrm{DPS}}$ distribution has been determined in *region II* and has been scaled to the detector level cross section of the $\Delta\mathrm{S}$ observable.

10.2.2 Migrations within the Phase Space

The first type of migrations are within the considered phase space. They correspond to events that pass the selection criteria on both generator and detector level, but of which the observables have significantly different values.

Depending on the unfolding procedure, the migrations within the phase space are studied in different ways. In what follows, it is opted to use an unfolding method based on the SVD. The method makes use of the matrix of migrations A, effectively connecting the detector and generator level. The migration matrices are expected to be diagonal whenever a strong correlation exists between the detector and the generator level. The largest values of the migration matrix are expected to be where the values of the detector and generator level observables are approximately the same.

Jet Matching Procedure

Before the proper migration matrices can be obtained, an additional spatial matching of the jets on detector and generator level needs to be performed. If one would naively fill the migration matrices without the matching, a significant background is found in all migration matrices. In Fig. 10.6, the migration matrices without any additional spatial matching for the pseudorapidity spectra are shown as an example. The large background finds it origin in the fact that the jets on detector and generator level are ordered in p_T. Due to the detector simulation, variations in the p_T of the jets can alter their order. The leading detector level jet does not necessarily correspond to the leading generator level jet. Both jets can be the leading jets in p_T but lie in vastly different regions of the detector. The background is created when measuring the observables for jets with different spatial coordinates at detector and generator level.

Through a spatial matching procedure of all jets on detector and generator level, the background can be removed from the migration matrices. The spatial matching criterion is given below.

$$R_{ij}^2 = \left(\eta_i^{\mathrm{gen}} - \eta_j^{\mathrm{det}}\right)^2 + \left(\phi_i^{\mathrm{gen}} - \phi_j^{\mathrm{det}}\right)^2 \leq R_\eta^2 + R_\phi^2 \tag{10.5}$$

Where the indices i and j iterate over the generator and detector level jets of the event. The cone radii R_η and R_ϕ are equal to 0.5. The radii are roughly ten times larger compared to the resolution in η and ϕ. Smaller radii would prohibit the matching of a fraction of corresponding detector and generator level jets due to resolution effects. While moving to larger cone radii would result the formation of a background. A detector level jet is matched to every generator level jet. If no match is found, the jet is omitted from the event. The jet with the smallest radius R_{ij} is taken as the match whenever multiple detector level jets match the generator level jet. Once a detector and generator level jet are matched, they are omitted from the rest of the matching procedure. The four leading jets on generator level, matched to a detector level jet, are taken as the input for all observables.

Migration Matrices of the Jet Matched Samples

The migration matrices for the p_T spectra, the η spectra and all observables are shown in Fig. 10.7, Fig. 10.8, Fig. 10.9 and Fig. 10.10 for the PYTHIA 8 sample, while Fig. 10.11,

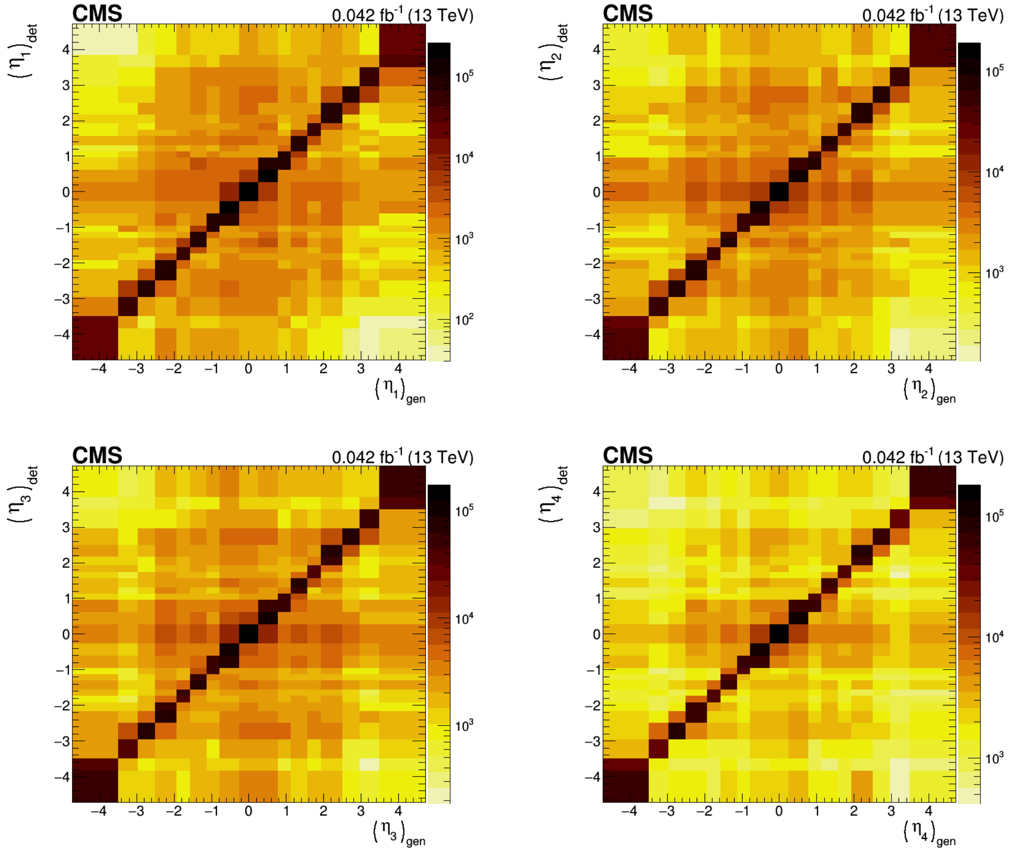


Figure 10.6: The migration matrices of the η spectra for the leading (top left), subleading (top right), third leading (bottom left) and fourth leading (bottom right) jets for the PYTHIA 8 MC sample without the spatial jet matching are shown.

Fig. 10.12, Fig. 10.13 and Fig. 10.14 depict the migration matrices for the HERWIG++ sample. The general shape of the migration matrices are comparable for both MC samples, apart from small bin-to-bin differences.

The p_{T} spectra show migrations between distant bins, however the off-diagonal bins rapidly decrease and are multiple orders lower compared to the bins on the diagonal. The pseudorapidity spectra of the four leading jets as well as the pseudorapidity spectra of the leading jet of the single inclusive jet processes show much fewer migrations compared to the p_{T} spectra. The background of the η spectra without any spatial matching, shown in Fig. 10.6, has been removed.

The angular variables $\Delta\phi_{\mathrm{soft}}$ and $\Delta\phi_{3\mathrm{j}}^{\mathrm{min}}$ show very few migrations, the value of off-diagonal bins immediately drops several orders, which can be attributed to the good resolution in ϕ. Diagonal behavior of the migration matrix for $\Delta\mathrm{Y}$ is also observed. It is a pseudorapidity-based observable and has a good resolution. The migration matrices for the p_{T}-based observables show wider spread migrations. The two softest jets enter the definition of $\Delta p_{\mathrm{T,soft}}$ in eq. (3.25), while all four jets enter in eq. (3.26) for the $\Delta\mathrm{S}$ observable. The bad resolution for low p_{T} jets make the observables highly sensitive to any differences in p_{T} and ϕ between detector and generator level, $\Delta\mathrm{S}$ even more so as $\Delta p_{\mathrm{T,soft}}$. The same non-diagonal behavior is observed in the migration matrices of the $\Delta\mathrm{S}_{\mathrm{DPS}}$ observable.

Although the spatial matching has been performed, the variable ϕ_{ij} shows a background throughout the phase space. The presence of the pedestal is due to the definition of the variable in eq. (3.24). It is the angular difference between the most forward and backward jet. Even though all four leading generator level jets have been matched to a detector level jet, the most forward (backward) jet on generator level does not always correspond the the most forward (backward) detector level jet. Therefore these most remote mismatched jets can lie in a completely different region of the azimuthal plane, effectively producing the background in the migration matrix of ϕ_{ij}. Any additional matching criteria, e.g. additional matching in p_{T}, would create a more artificial migration matrix.

The importance of the model reweighing and the spatial matching can be highlighted here. If the reweighing procedure is not performed, the bin-by-bin differences in the migration matrices can differ, leading to an overcorrection or undercorrection of the data in the unfolding procedure. The removal of the background through the spatial matching procedure, apart from the ϕ_{ij} observable, will be beneficial for the stability of the unfolding procedure.

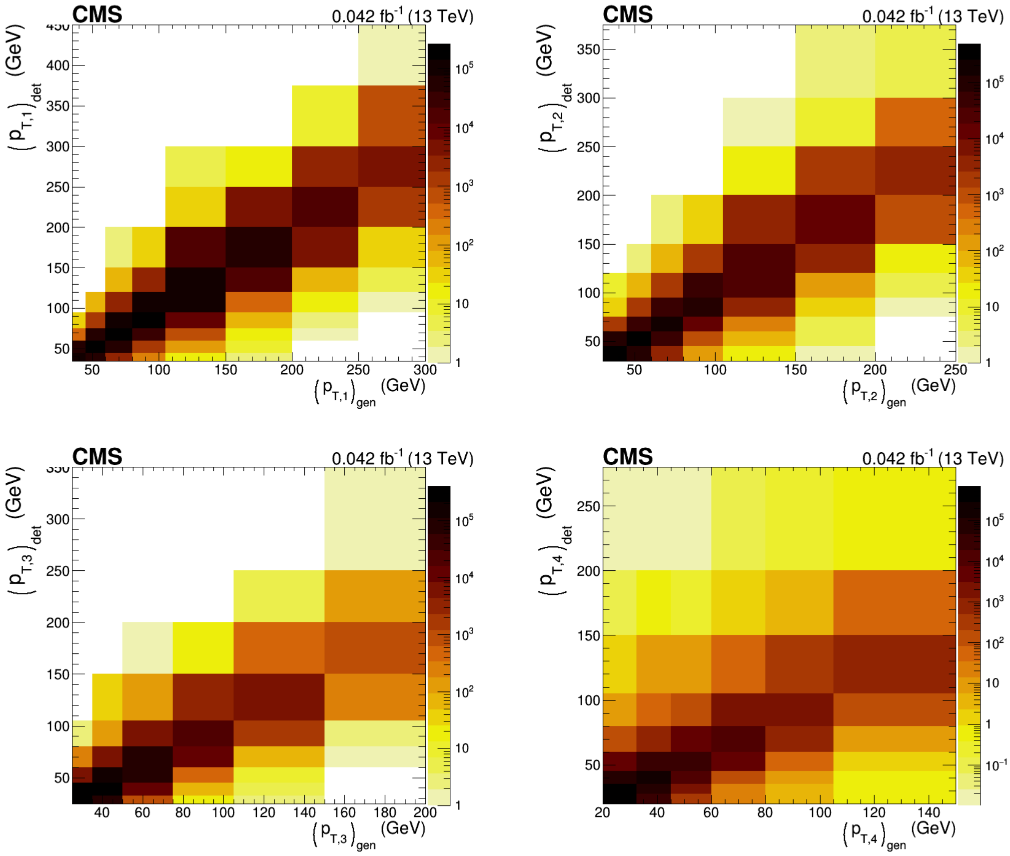


Figure 10.7: The migration matrices of the p_{T} spectra for the leading (top left), subleading (top right), third leading (bottom left) and fourth leading (bottom right) jets for the PYTHIA 8 MC sample are shown.

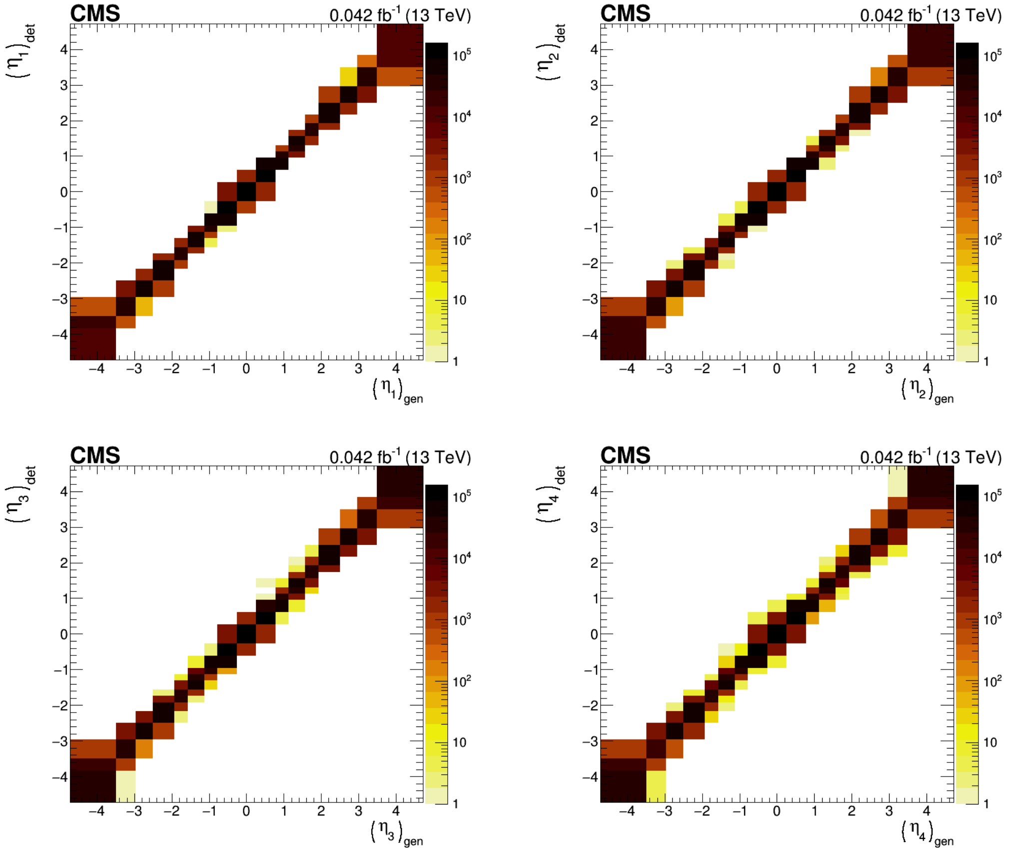


Figure 10.8: The migration matrices of the η spectra for the leading (top left), subleading (top right), third leading (bottom left) and fourth leading (bottom right) jets for the PYTHIA 8 MC sample are shown.

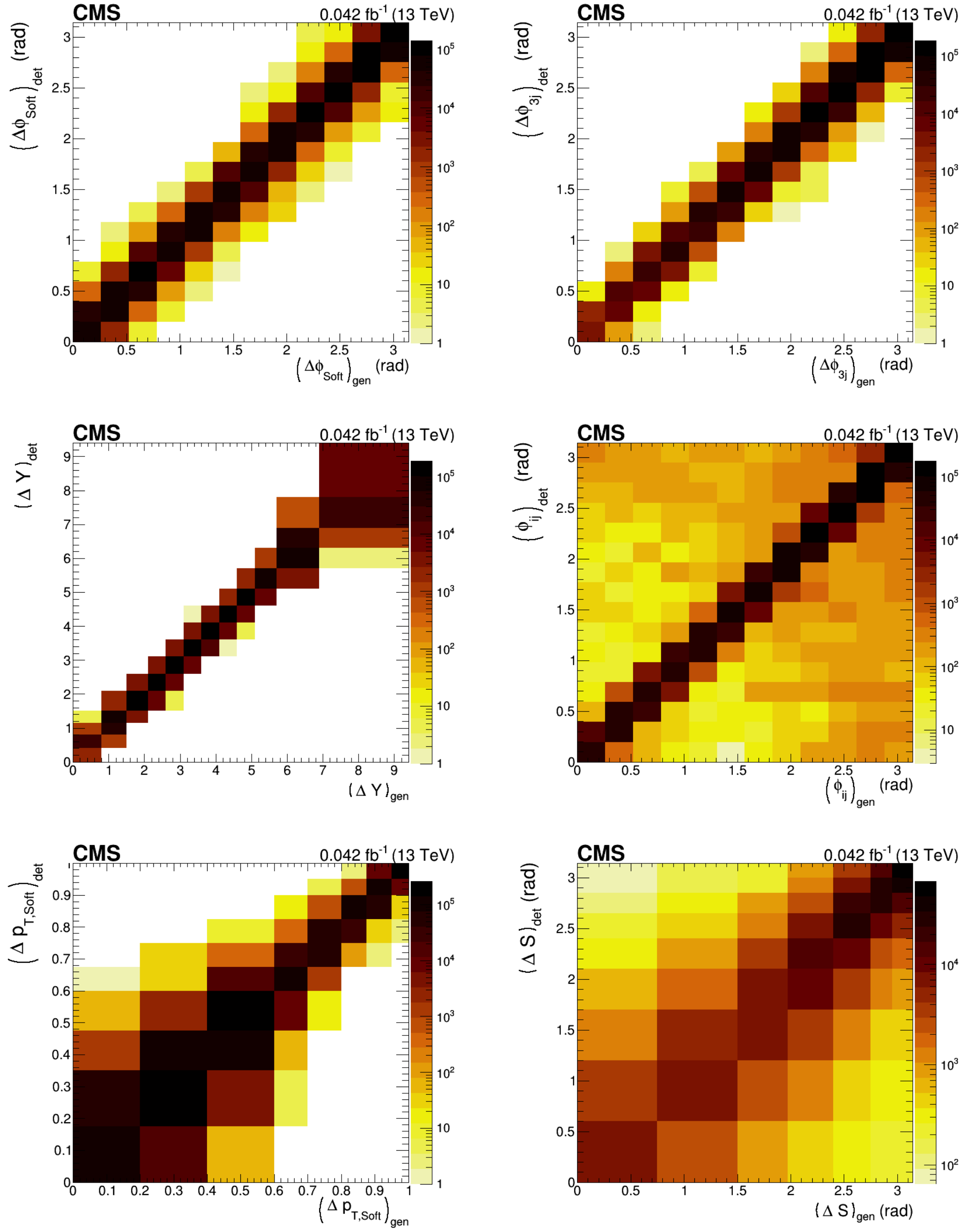


Figure 10.9: The migration matrices from the PYTHIA 8 sample for the DPS-sensitive observables are shown here.

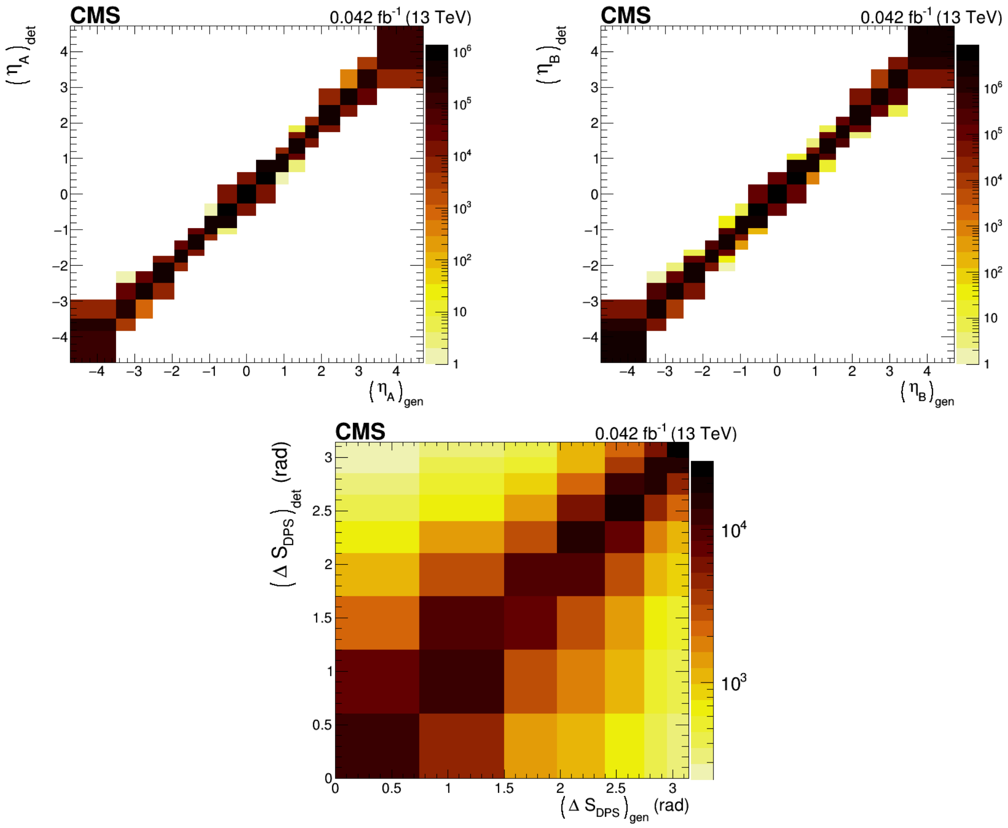


Figure 10.10: The migration matrices for the pseudorapidity spectra of the inclusive single jet cross sections, obtained from the PYTHIA 8 sample, are shown on the top left and top right plots. The migration matrix for the ΔS_{DPS} observable from the pure DPS PYTHIA 8 sample is depicted on the bottom plot.

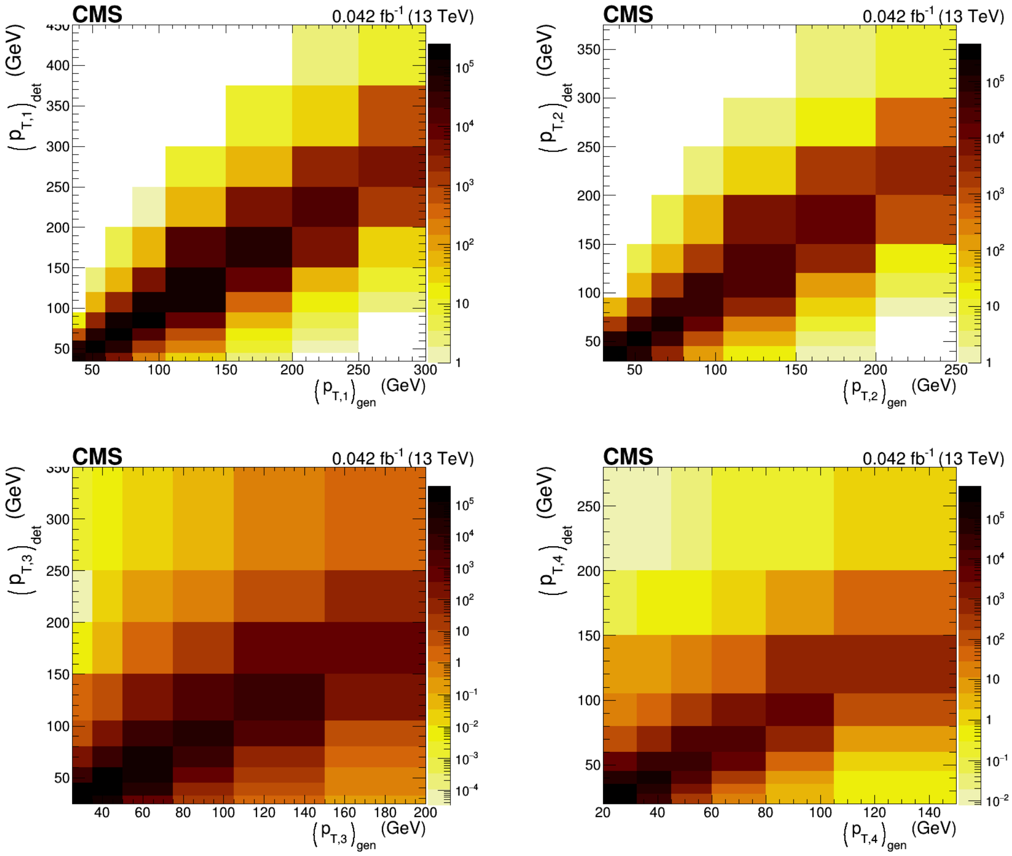


Figure 10.11: The migration matrices of the p_{T} spectra for the leading (top left), subleading (top right), third leading (bottom left) and fourth leading (bottom right) jets for the HERWIG++ MC sample are shown.

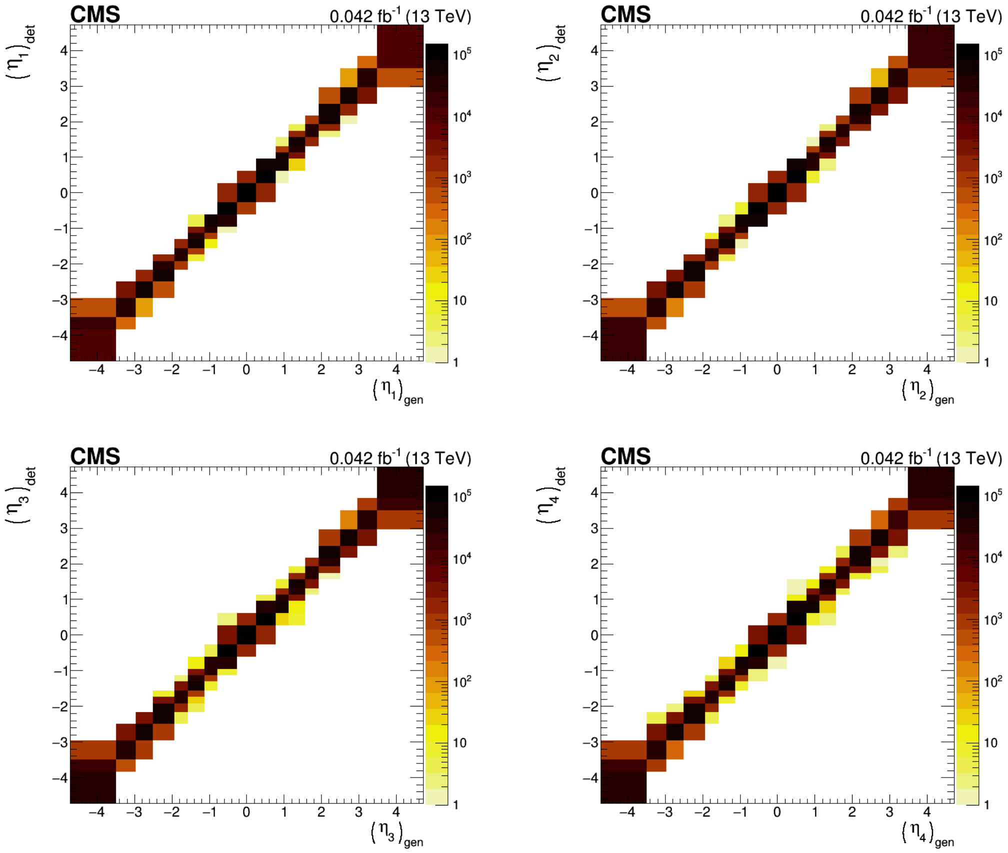


Figure 10.12: The migration matrices of the η spectra for the leading (top left), subleading (top right), third leading (bottom left) and fourth leading (bottom right) jets for the HERWIG++ MC sample are shown.

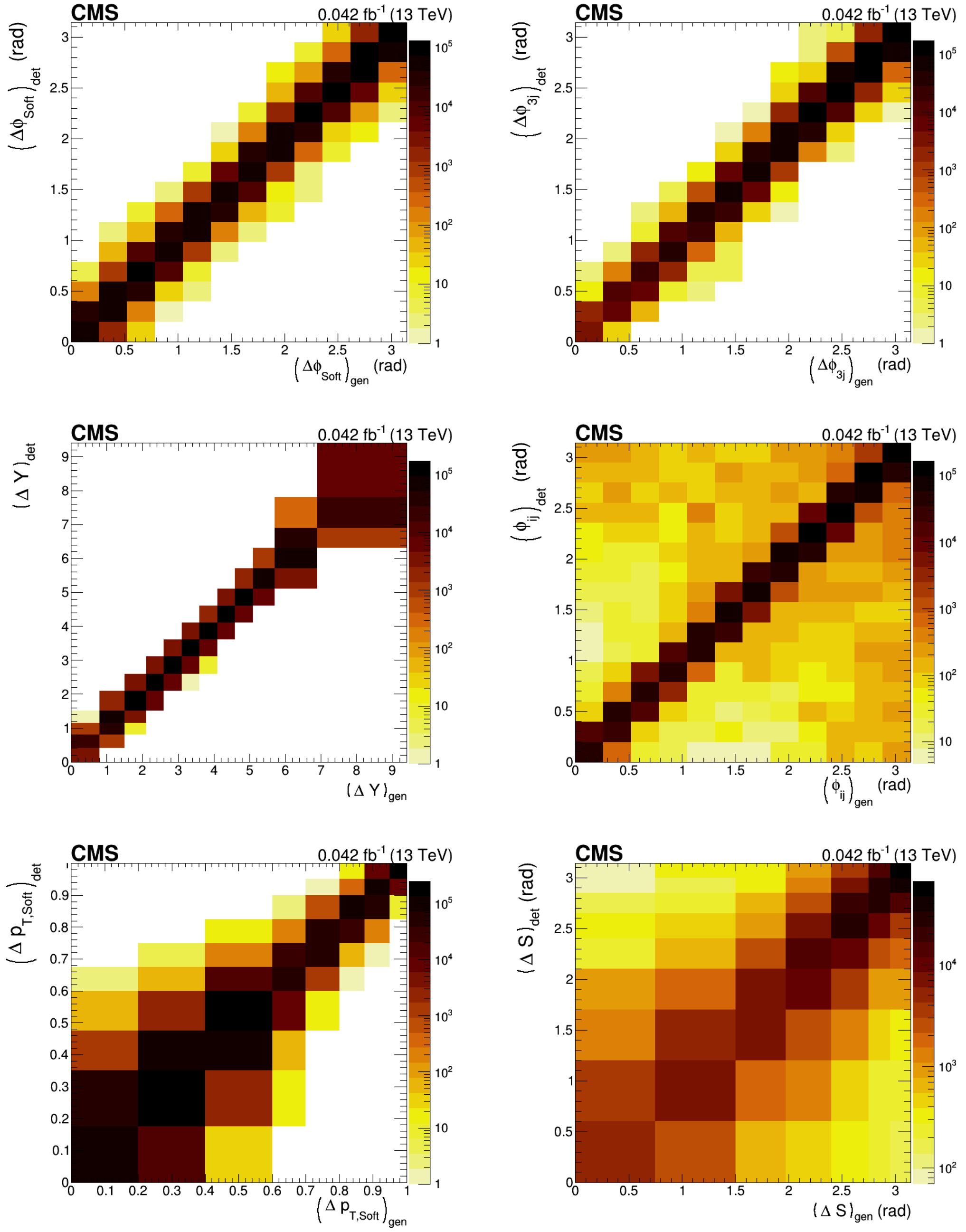


Figure 10.13: The migration matrices from the HERWIG++ sample for the DPS-sensitive observables are shown here.

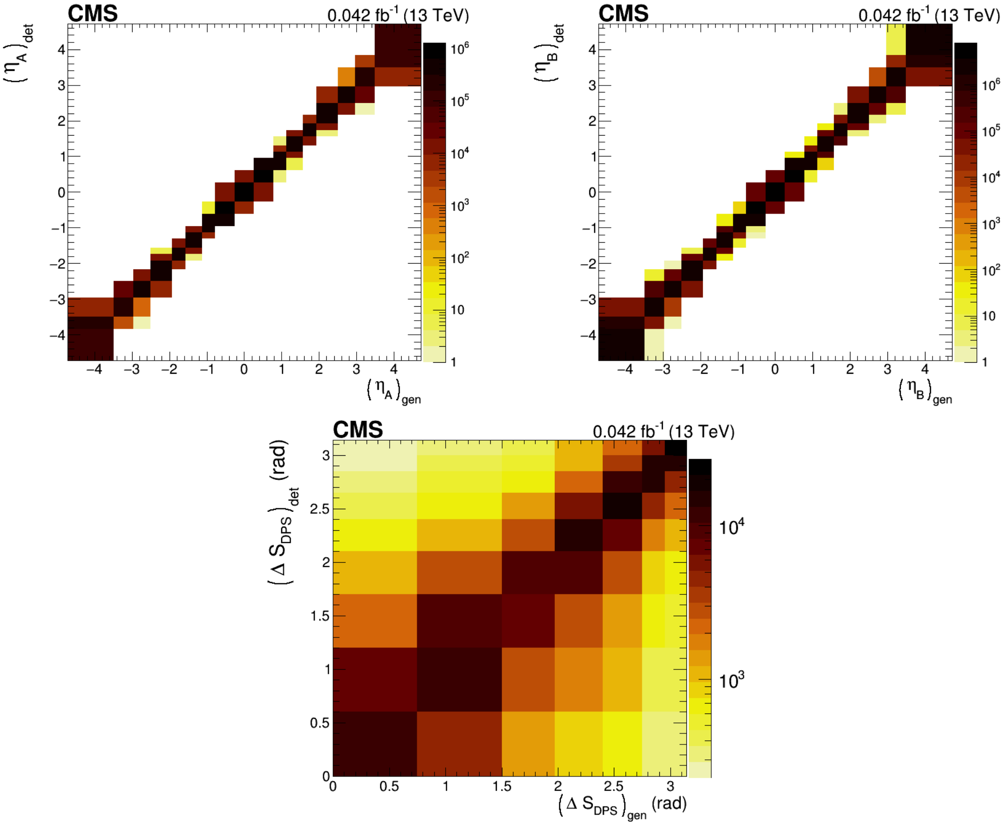


Figure 10.14: The migration matrices for the pseudorapidity spectra of the inclusive single jet cross sections, obtained from the HERWIG++ sample, are shown on the top left and top right plot. The migration matrix for the ΔS_{DPS} observable from the pure DPS HERWIG++ sample is depicted on the bottom plot.

10.2.3 Migrations into and out of the Phase Space

The second type of migrations are into or out of the considered phase space. Meaning that the events either pass the selection criteria on detector or generator level, but not the criteria on both levels. The migrations are labeled as fakes and misses and are defined below.

- The fake rate of an observable measures the ratio of events in the i'th detector level bin, which are selected only at detector level, compared to all events that pass the detector level.

  $$F_i = \frac{N_{\text{only det}}(O^{\text{det}} \in \text{bin}_i)}{N_{\text{det}}(O^{\text{det}} \in \text{bin}_i)} \tag{10.6}$$

  Where O represents the observable in question. The subscript “det” indicates the more inclusive selection. Where the subscript “only det” refers to events that pass the detector level selection but not the generator level criteria.

- The miss rate is the same ratio as the fakes but defined at generator level, yielding:

  $$M_i = \frac{N_{\text{only gen}}(O^{\text{gen}} \in \text{bin}_i)}{N_{\text{gen}}(O^{\text{gen}} \in \text{bin}_i)} \tag{10.7}$$

  The subscripts are defined analogue to the subscripts in the definition of the fake rate.

From the definitions of the fake and miss rates, it should be clear that the fake (miss) rate will use the detector (generator) level binning.

The fakes and misses need to be accounted for in the whole unfolding procedure. The fakes are detector level events or measurements with no corresponding generator level event or truth entry. The fakes are essentially causeless-effects and correspond to the background vector **b** from the formal discussion of the unfolding procedure in Sec. 10.1. The misses are the opposite of fakes, i.e. truth entries with no corresponding measurement. They can be viewed as inefficiencies in the simulation procedure and are in essence effectless-causes. The fakes will be subtracted from and relative to the detector level distribution of the observables obtained from the data sample. The misses, however, will be scaled and added to the unfolded distribution of the observables, in order to compensate for the inefficiencies of the simulation procedure.

The fake and miss rate for all observables is shown in Fig. 10.15, Fig. 10.16, Fig. 10.17 and Fig. 10.18 for the PYTHIA 8 sample. While Fig. 10.19, Fig. 10.20, Fig. 10.21 and Fig. 10.22 show the same distributions for the HERWIG++ sample.

The fake and miss rate for all observables, apart from the p_{T} spectra, are distributed quite uniformly. For the fake (miss) rate, it is observed that the probability that a generator level jet below (above) the p_{T} threshold is simulated at detector level above (below) the p_{T} threshold decreases as the p_{T} of the detector (generator) level jet increases. The result is a steep drop of the fake and miss rate when moving away from the p_{T} thresholds in the p_{T} spectra. The fraction of the fakes and misses for the p_{T} spectra of the jets starts around 70% for the lowest bins in p_{T} and diminishes quickly to values smaller than 20%. Apart from the p_{T} spectra and the $\Delta\mathrm{S}_{\mathrm{DPS}}$ observable, the fake rate ranges from 20% to 55%, while the miss rate lies between 30% and 60%. The miss rate is in general larger than the fake rate. The difference between the two rates is overall larger for the PYTHIA 8 sample compared to the HERWIG++ sample.

For the inclusive single jet processes, the fake and miss rate as a function of the p_{T} of the jets is also presented in Fig. 10.18 and Fig. 10.22. From these plots, it becomes clear that the size of the fake and miss rate observed in the η spectra is due the same boundary effect as observed in the fake and miss rate of the inclusive four jet production in Fig. 10.15 and Fig. 10.19. The fake rate ranges from 55% at low values of p_{T} down to a negligible amount at large transverse momenta, while the miss rate starts at 40% and drops down to 5%. For the p_{T} spectra, a similar binning compared to the binning of the four leading jets has been used, however, the spectra are only introduced here to show the origin of the fake and miss rate. They will not be unfolded or used elsewhere in the analysis.

A larger fake and miss rate is observed for the $\Delta\mathrm{S}_{\mathrm{DPS}}$ observable. The fake and miss rate varies between 55% and 80%. The high fake and miss rate compared to all other observables can be understood from the mixing procedure. Two separate inclusive single jet events are combined into an inclusive four jet event. Whenever one of the two or both events are a fake (miss), the combined event will also be a fake (miss), effectively increasing the probability of a pure DPS event to be a fake (miss).

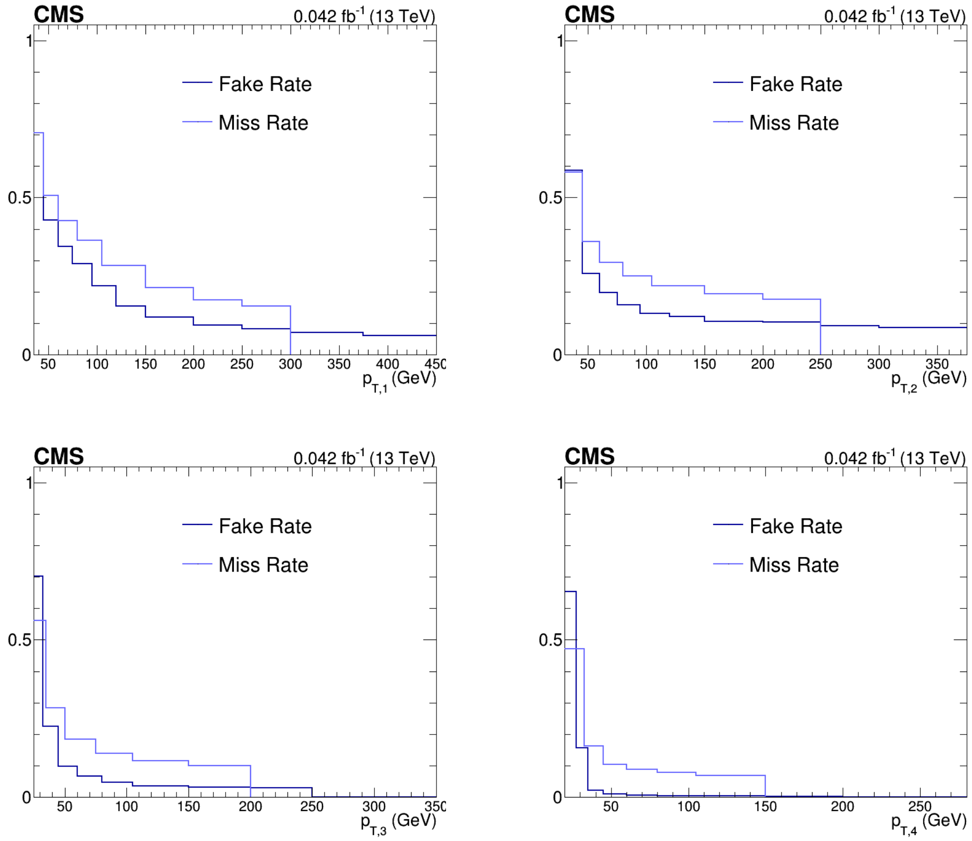


Figure 10.15: The fake and miss rate for the p_{T} spectra of the leading (top left), subleading (top right), third leading (bottom left) and fourth leading (bottom right) jets for the PYTHIA 8 MC sample are shown.

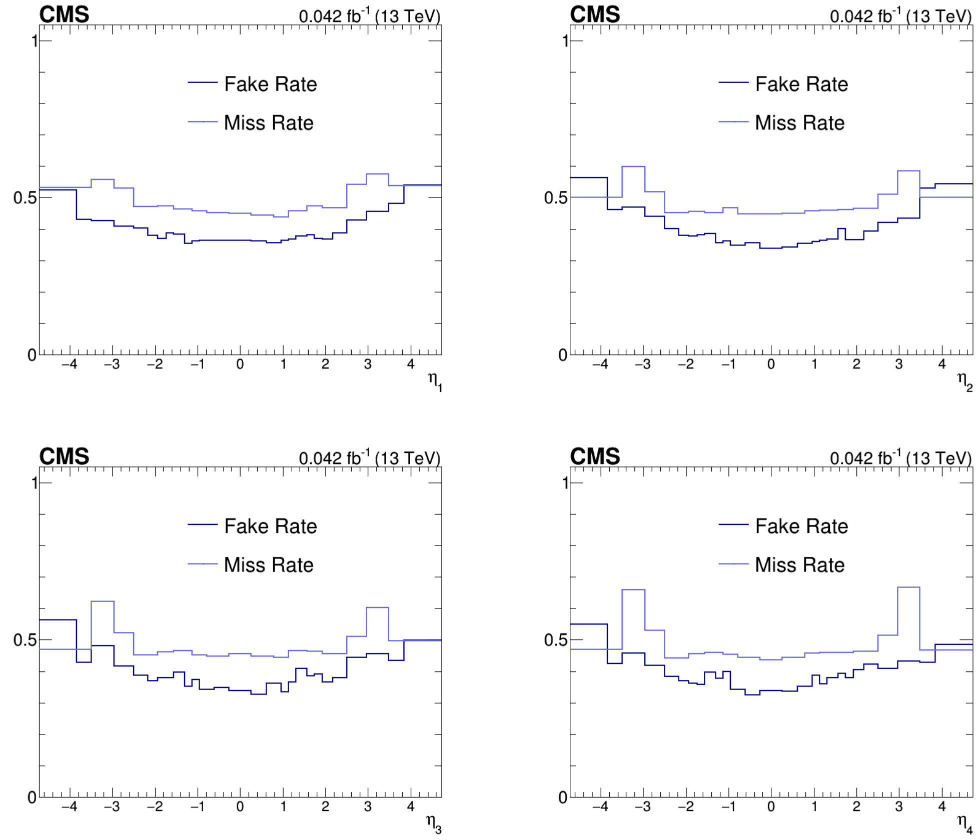


Figure 10.16: The fake and miss rate for the η spectra of the leading (top left), subleading (top right), third leading (bottom left) and fourth leading (bottom right) jets for the PYTHIA 8 MC sample are shown.

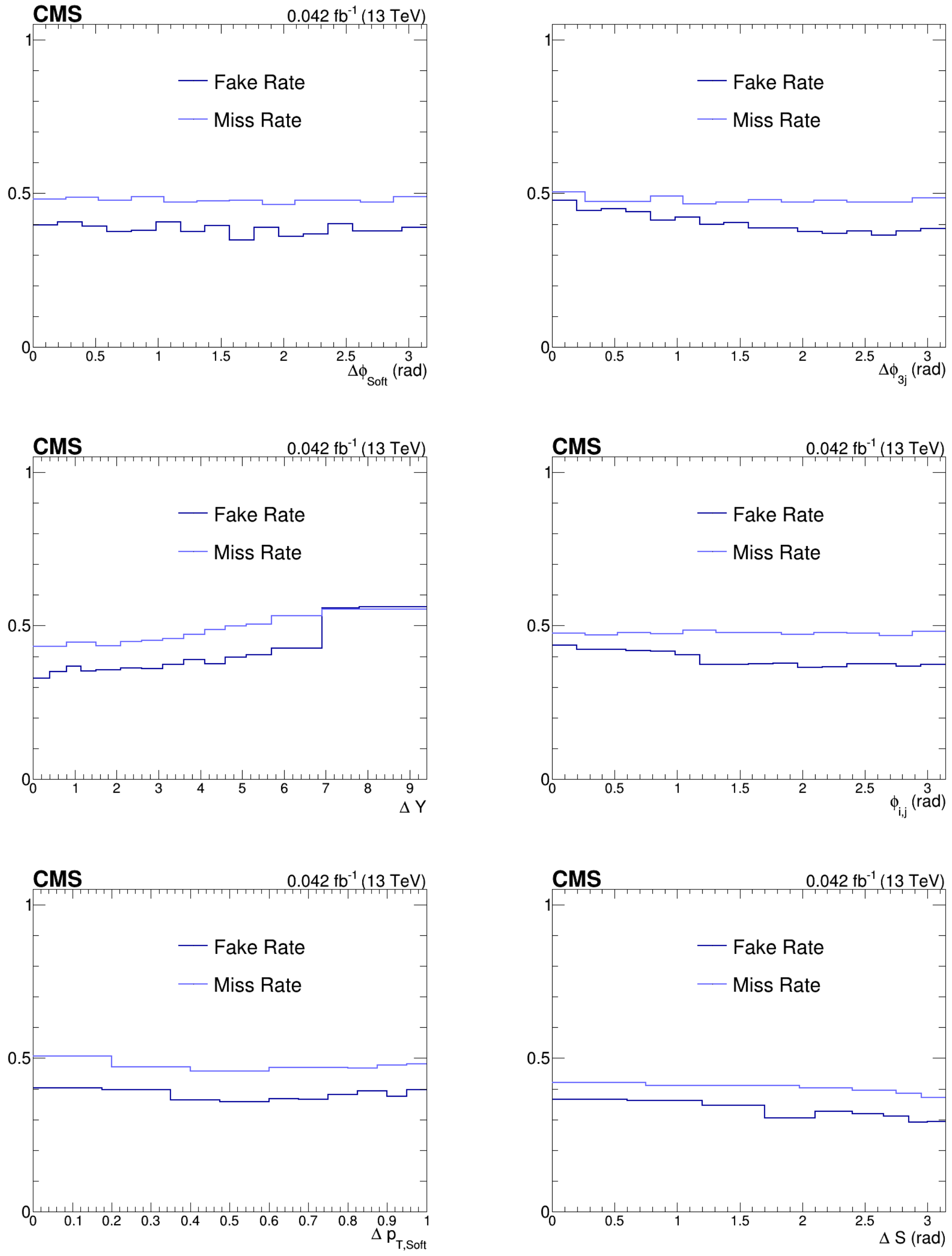


Figure 10.17: The fake and miss rate for the DPS-sensitive observables for the PYTHIA 8 sample. The ΔS observable (bottom right) has been determined in *region II*, compared to *region I* for all other observables.

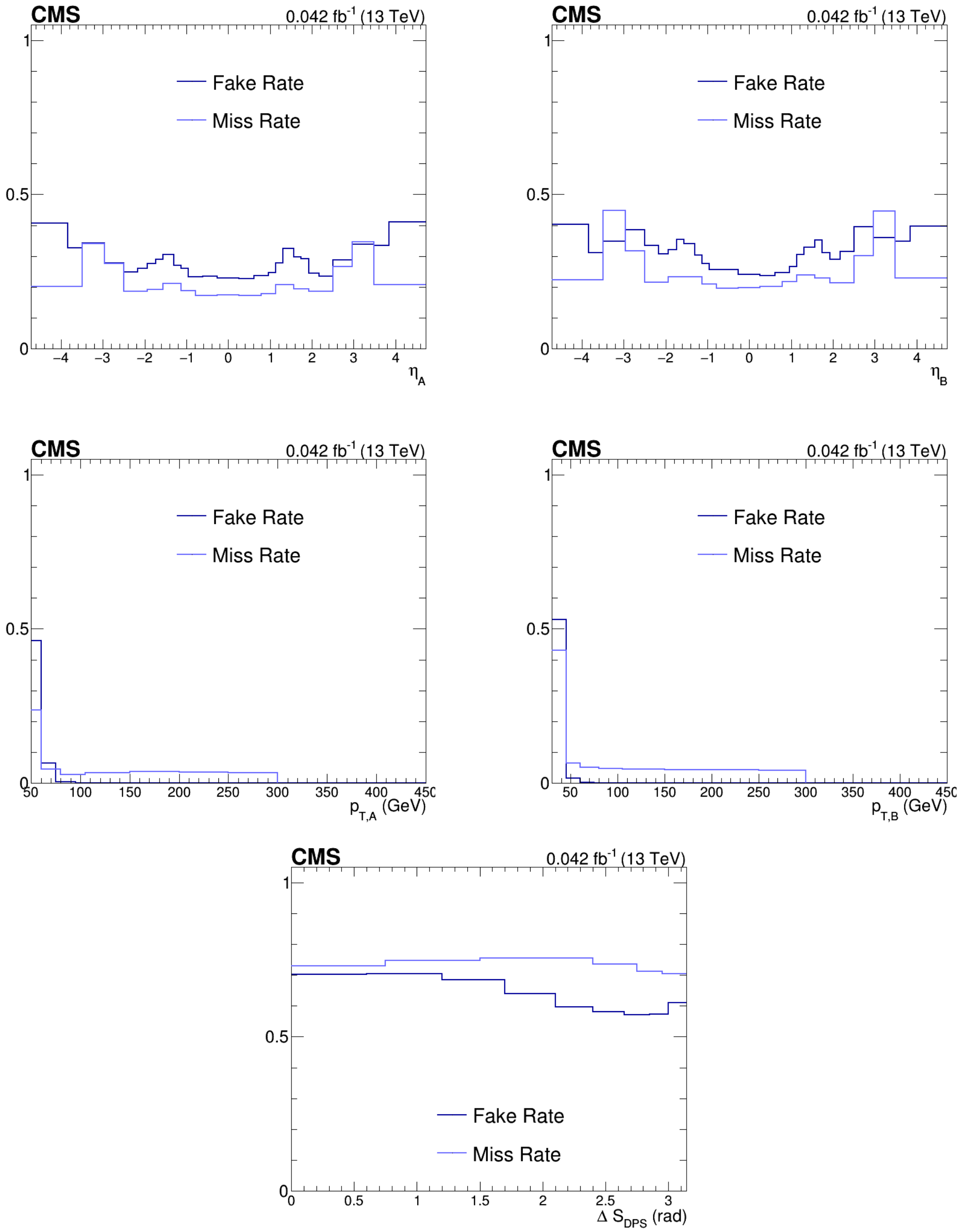


Figure 10.18: The fake and miss rate for the η and p_T spectra of the inclusive single jet cross sections for the PYTHIA 8 MC sample together with for the ΔS_{DPS} observable determined for the pure DPS PYTHIA 8 sample.

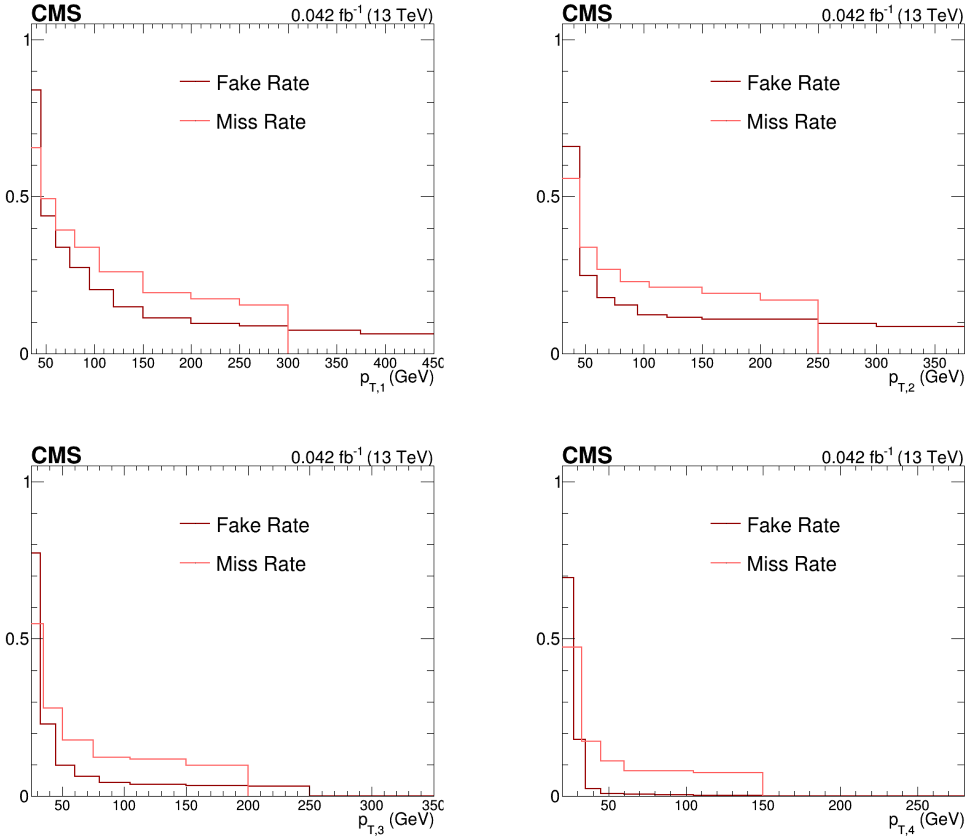


Figure 10.19: The fraction of fakes and misses for the p_{T} spectra of the leading (top left), sub-leading (top right), third leading (bottom left) and fourth leading (bottom right) jets for the HERWIG++ MC sample are shown.

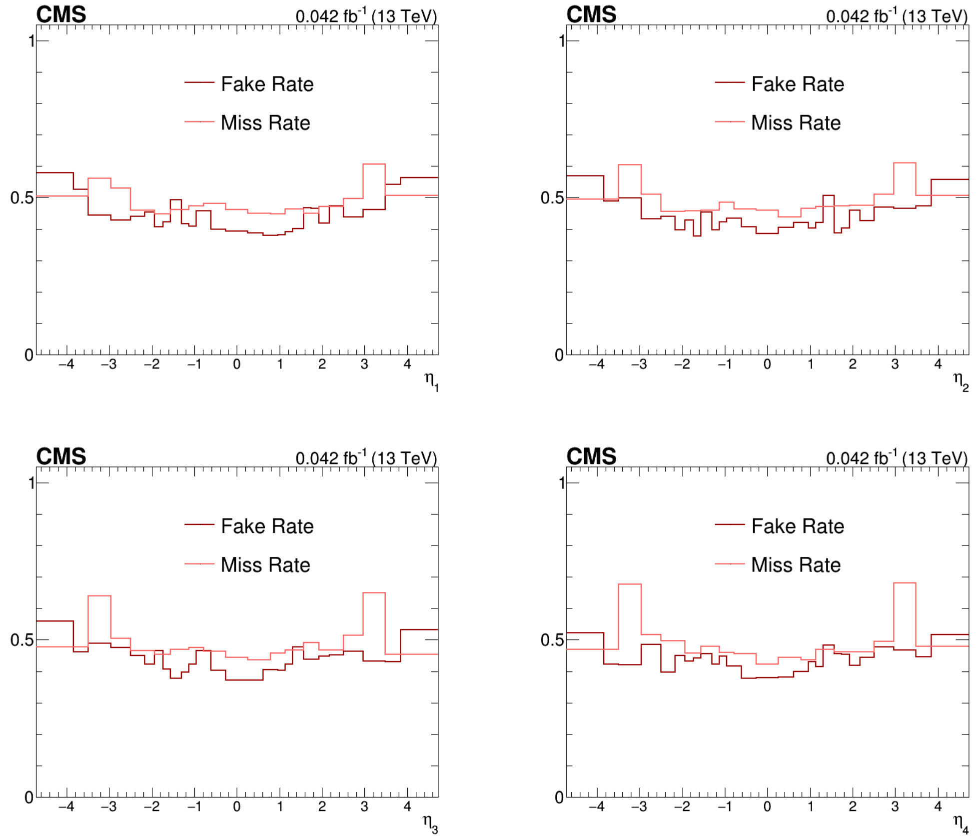


Figure 10.20: The fraction of fakes and misses for the η spectra of the leading (top left), sub-leading (top right), third leading (bottom left) and fourth leading (bottom right) jets for the HERWIG++ MC sample are shown.

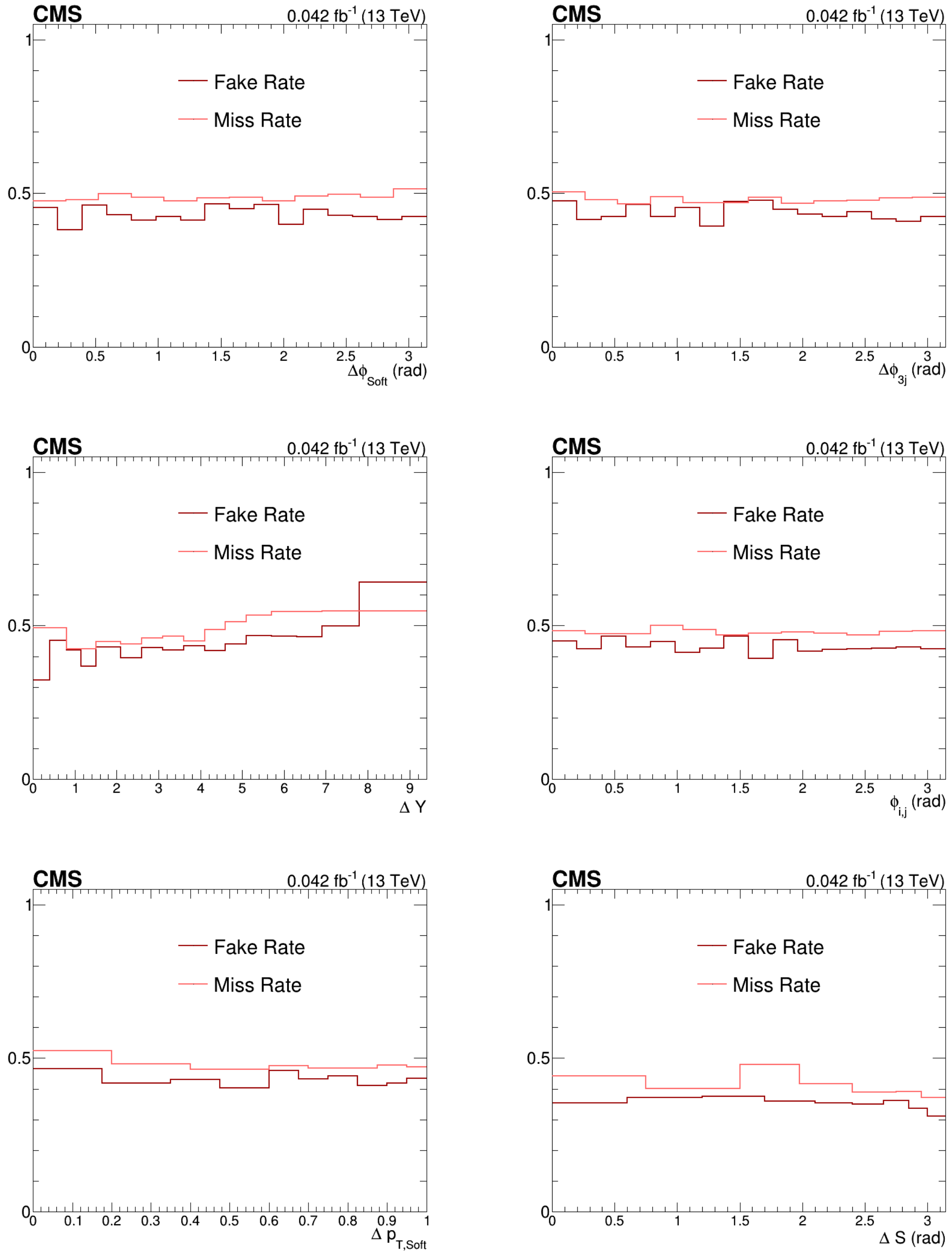


Figure 10.21: The fake and miss rate for the DPS-sensitive observables for the HERWIG++ sample. The ΔS observable (bottom right) has been determined in *region II*, compared to *region I* for all other observables.

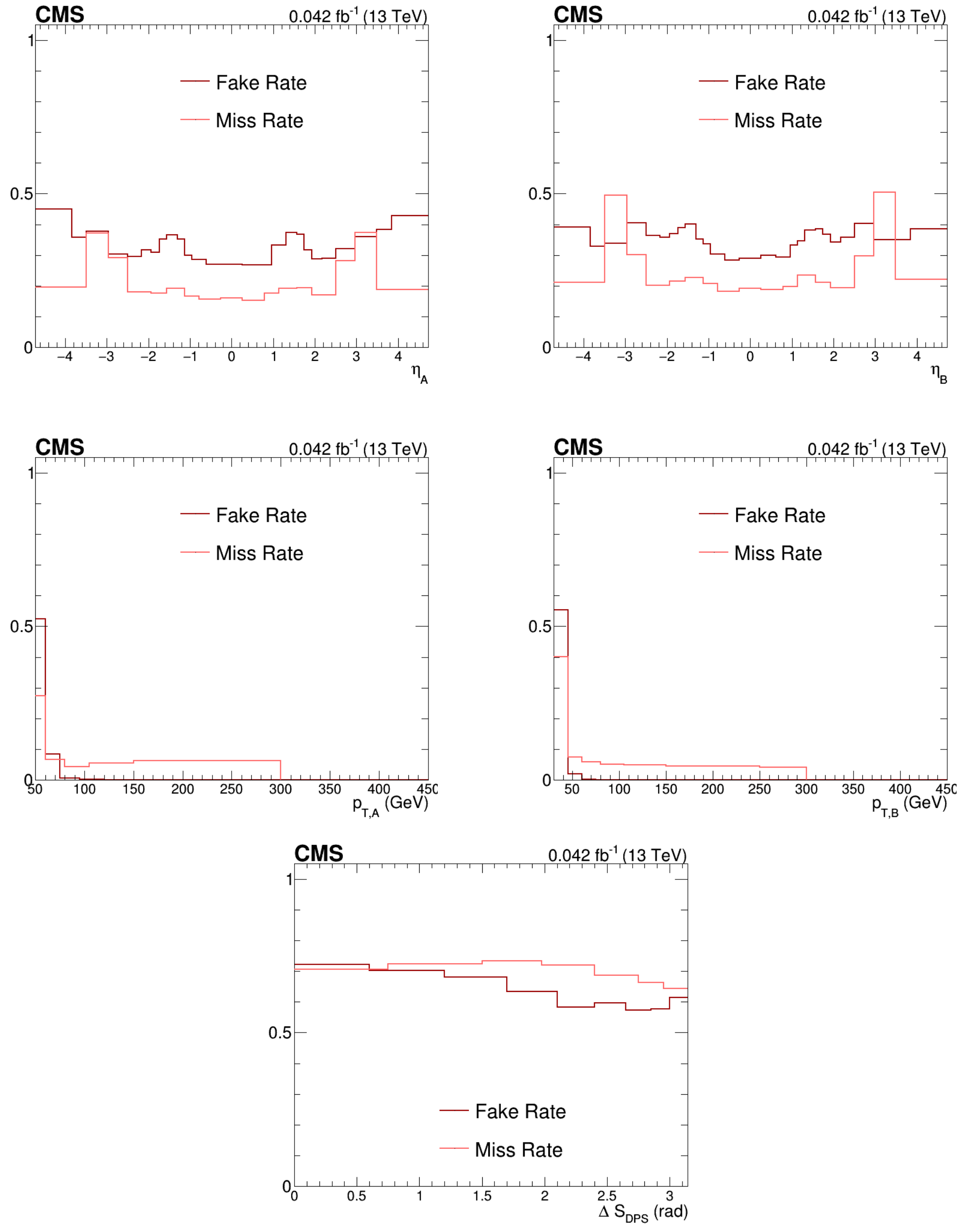


Figure 10.22: The fake and miss rate for the η and p_T spectra of the inclusive single jet cross sections for the HERWIG++ MC sample together with for the ΔS_{DPS} observable determined for the pure DPS HERWIG++ sample.

10.3 The `TUnfold` Algorithm

The `TUnfold` algorithm [179–181] has been used for the unfolding of all the observables and has implemented a least-squares minimization with Tikhonov regularization [178]. The stationary point or the minimum of the Lagrangian given below, is determined and will give the corresponding optimal estimator of the truth vector.

$$\mathcal{L}(x,\lambda) = \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3, \quad \text{with} \tag{10.8}$$

$$\mathcal{L}_1 = (\mathbf{y} - \mathbf{A}\mathbf{x})^{\mathrm{T}} \mathbf{V_y}^{-1} (\mathbf{y} - \mathbf{A}\mathbf{x}), \tag{10.9}$$

$$\mathcal{L}_2 = \tau^2 (\mathbf{x} - f_b \mathbf{x_0})^{\mathrm{T}} \left(\mathbf{L}^{\mathrm{T}}\mathbf{L}\right) (\mathbf{x} - f_b \mathbf{x_0}), \tag{10.10}$$

$$\mathcal{L}_3 = \lambda \left(Y - \mathbf{e}^{\mathrm{T}}\mathbf{x}\right), \tag{10.11}$$

$$\text{where} \quad Y = \sum_i y_i \quad \text{and} \quad e_j = \sum_i A_{ij} \tag{10.12}$$

The first term $\mathcal{L}_1$ of the full Lagrangian originates from the least-squares minimization, where

- The vector $\mathbf{y}$ has n rows and corresponds to the measurement.
- $\mathbf{V_y}$ is the covariance matrix of $\mathbf{y}$.
- The matrix $\mathbf{A}$ is the matrix describing the migrations which is obtained from the migration matrix $\mathbf{M}$ that in turn contains the total events counts from the MC sample used in the unfolding procedure. The matrix $\mathbf{A}$ is normalized to its rows or

  $$A_{ij} = \frac{M_{ij}}{\sum_{i=0}^{n} M_{ij}} \tag{10.13}$$

  The elements A_{ij} can now be interpreted as the probability of an observable in bin i at generator level migrating to a bin j at detector level.
- The vector $\mathbf{x}$ is the estimator of the truth vector $\tilde{\mathbf{x}}$, which is returned by the algorithm.

One assumes that $\mathbf{y}$ obeys a multivariate Gaussian distribution, where the mean $\tilde{\mathbf{y}}$ and the set of truth parameters $\tilde{\mathbf{x}}$, obey $\tilde{\mathbf{y}} = \mathbf{A}\tilde{\mathbf{x}}$.

The second part $\mathcal{L}_2$ of the full Lagrangian describes the regularization. Imposing regularization conditions is necessary to dampen the statistical fluctuations of $\mathbf{y}$ since they are amplified in the calculation of $\mathbf{x}$. In the Lagrangian L_2 one defines the parameters as below.

- The parameter τ is the regularization strength. The stationary point or minimum of the whole Lagrangian is determined for a fixed value of τ.
- The matrix $\mathbf{L}$ is the $n \times n_R$ regularization matrix, where n_R is the amount of rows and is equal to the number of regularization conditions.

- The bias vector $f_b\mathbf{x_0}$ is the product of a normalization factor f_b and a vector $\mathbf{x_0}$ defined as

$$(\mathbf{x_0})_j = \sum_{i=0}^{n} M_{ij} \tag{10.14}$$

  The role of $f_b\mathbf{x_0}$ is to suppress large deviations of the best estimator $\mathbf{x}$ from $f_b\mathbf{x_0}$, where x_0 is a generator level distribution determined by the MC sample.

Three different types of regularization are possible: regularization by the size of $\mathbf{x}$, by the first derivative of $\mathbf{x}$ and by the second derivative of $\mathbf{x}$ or the curvature of $\mathbf{x}$. Each type of regularization corresponds to a different regularization matrix $\mathbf{L}$. A complete discussion on these different regularization methods can be found in [179].

The third and last Lagrangian $\mathcal{L}_3$ is an area constraint. The presence of the Lagrangian multiplier λ ensures that the integral of optimal estimator $\mathbf{x}$ is equal to the integral of the input distribution $\mathbf{y}$. In the `TUnfold` algorithm, the inclusion of the third Lagrangian is optional and can be left out of the whole unfolding procedure.

10.4 The Regularization Strength

The unfolding algorithm determines the best estimator $\mathbf{x}$ given a certain value of τ. However, the size of the regularization strength can impact the result heavily. A value that is too small or too large can lead to overfitting or underfitting, respectively, which is schematically depicted in Fig. 10.23. The optimal value of τ is determined in an iterative procedure, the so-called L-curve scan. The two functions below are considered in the procedure.

$$L_x^{\text{curve}} = \log \mathcal{L}_1 \quad \text{and} \quad L_y^{\text{curve}} = \log \frac{\mathcal{L}_2}{\tau^2} \tag{10.15}$$

If the regularization strength is too small or $\tau^2 \to 0$, then L_x^{curve} will be minimal and L_y^{curve} is maximal. The total Lagrangian in eq. (10.8) will be dominated by $\mathcal{L}_1$ since $\mathcal{L}_2$ is negligible. The estimator $\mathbf{x}$ will not be affected by the regularization and will lie close to the exact solution of inverted problem in eq. (10.1) or

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} \tag{10.16}$$

Where $\mathbf{y}$ is the measured distribution. As a result none of the possible fluctuations have been dampened, leaving room for the possibility of overfitting. The other extreme is where τ^2 is too large. The total Lagrangian in eq. (10.8) will be dominated by the regularization term $\mathcal{L}_2$. Fluctuation will be dampened, but often too hard, resulting in an underfitting of the data. The optimal value of τ^2 lies in between these two cases and is situated where the curvature of both functions is maximal. The curvature of the L-curve in the `TUnfold` package is determined by the derivative given below.

$$\mathcal{C} = \frac{\mathrm{d}L_x^{\mathrm{curve}}\mathrm{d}^2L_y^{\mathrm{curve}} - \mathrm{d}^2L_x^{\mathrm{curve}}\mathrm{d}L_y^{\mathrm{curve}}}{\left((\mathrm{d}L_x^{\mathrm{curve}})^2 + (\mathrm{d}L_y^{\mathrm{curve}})^2\right)^{\frac{3}{2}}} \tag{10.17}$$

In an iterative procedure the unfolding is repeated for a range of values of $\log\tau$. The curvature is determined for each point and is parametrized by a cubic spline. The point yielding the the largest curvature corresponds to the approximated value of the optimal regularization strength τ^2.

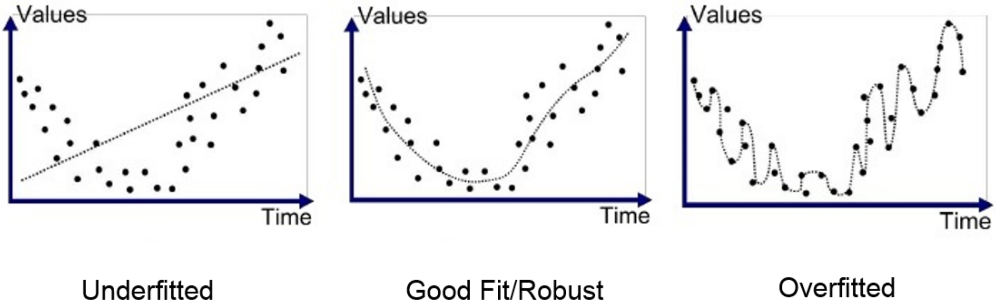


Figure 10.23: Illustration of the importance of the size of the regularization strength. A non-optimal choice can lead to overfitting or underfitting of the problem, as depicted on the left and right plot, respectively.

10.5 Unfolding of the Data

Two free parameters remain in the whole unfolding procedure: the type of regularization and the use of the additional area constraint. In an effort to determine the optimal combination of the free parameters, the unfolding is repeated for all possible choices of the two parameters. The optimal choice is determined by folding the distributions back to detector level and by minimizing the the corresponding χ^2-test. The χ^2-test compares the original detector level distribution with the backfolded distribution, where χ^2_{backf} is defined below.

$$\chi^2_{\mathrm{backf}} = \sum_{i=1}^{n}\left(\frac{y_i - y_i^{\mathrm{backf}}}{\sqrt{\sigma_i^2 + (\sigma_i^{\mathrm{backf}})^2}}\right)^2, \tag{10.18}$$

$$\text{where} \quad y_i^{\mathrm{backf}} = \sum_{j=1}^{m} A_{ij}x_j + b_i \tag{10.19}$$

Where x_j is the best estimator of the true distribution $\tilde{x}_j$, determined by the L-curve scan. Due to substantial regularization, the NDF can differ from the number of bins.

The true NDF is given by the rank of the matrix $\mathbf{A}\mathbf{V_y}\mathbf{V_y}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$. The reduced χ^2_{backf}, i.e. $\chi^2_{\mathrm{backf}}/\mathrm{NDF}$, is minimized. The procedure is repeated for all observables individually.

The whole unfolding procedure, along with the determination of the optimal configuration, entails the following steps:

- The fake rate is handled as a background and is subtracted from the detector distribution. The `TUnfold` package subtracts and deals with the error propagation accordingly.
- An iterative procedure entailing the L-curve scan is performed. The optimal regularization strength τ^2 and the corresponding unfolded distribution are determined.
- The miss rate is taken into account by a bin-by-bin addition relative to the unfolded spectrum.
- The first three steps are repeated for different regularization conditions and with the area constraint turned on and of. For each combination, the corresponding $\chi^2_{\mathrm{backf}}/\mathrm{NDF}$ is calculated.
- The unfolded distribution with a minimal $\chi^2_{\mathrm{backf}}/\mathrm{NDF}$ is chosen as the optimal unfolded distribution.
- All the previous steps are repeated. The PYTHIA 8 and HERWIG++ sample are both used for the unfolding. Both sets of optimal unfolded curves are kept.
- The average of the observables unfolded by both the PYTHIA 8 and HERWIG++ sample is taken as the final unfolded curve.

Since the unfolding procedure relies on a MC model determining the migrations, the results will be model dependent. The two final steps are performed in an effort to reduce the model dependence and the bias towards a certain model. The usage of two different models will be addressed again when discussing the systematical errors of the whole analysis in Chap. 11.

The unfolded distributions of the p_{T} spectra, the pseudorapidity spectra and the different observables are given in Fig. 10.24, Fig. 10.25, Fig. 10.26 and Fig. 10.27, respectively. The observables unfolded with both MC samples are given as well as the average of the two distributions. The difference between the observables unfolded with PYTHIA 8 and HERWIG++ is at most 10%, except for the first bin of the p_{T} spectrum of the leading jet.

The backfolded distributions for data unfolded with PYTHIA 8 and HERWIG++ are shown in Fig. 10.28, Fig. 10.29, Fig. 10.30 and Fig. 10.31. The p_{T} spectra are only folded back for the generator level range since the information of the higher p_{T} bins is lost in the unfolding procedure. All backfolded distributions are in agreement with the original detector level distributions within their respective statistical uncertainty for almost all bins. Only the second bin for the p_{T} spectrum of the last leading jet shows a deviation.

The exact values of the minimal $\chi^2_{\mathrm{backf}}/\mathrm{NDF}$, the regularization condition and whether or not the additional area constraint is used, are given in Tab. 10.1 for the PYTHIA 8 and HERWIG++ samples, along with the condition number which never exceeds 10^2. Therefore, the unfolding of all observables is a well-posed problem and only little regularization is needed in the procedure itself. All observables yield the lowest values for $\chi^2_{\mathrm{backf}}/\mathrm{NDF}$

with the size regularization condition and with the additional area constraint turned on, except the p_{T} spectrum of the second leading jet when unfolded with the PYTHIA 8 sample. In order to obtain a minimal $\chi^2_{\mathrm{backf}}/\mathrm{NDF}$, the regularization by the first derivative and no extra area constraint is needed in the unfolding procedure for the p_{T} spectrum of the second leading jet.

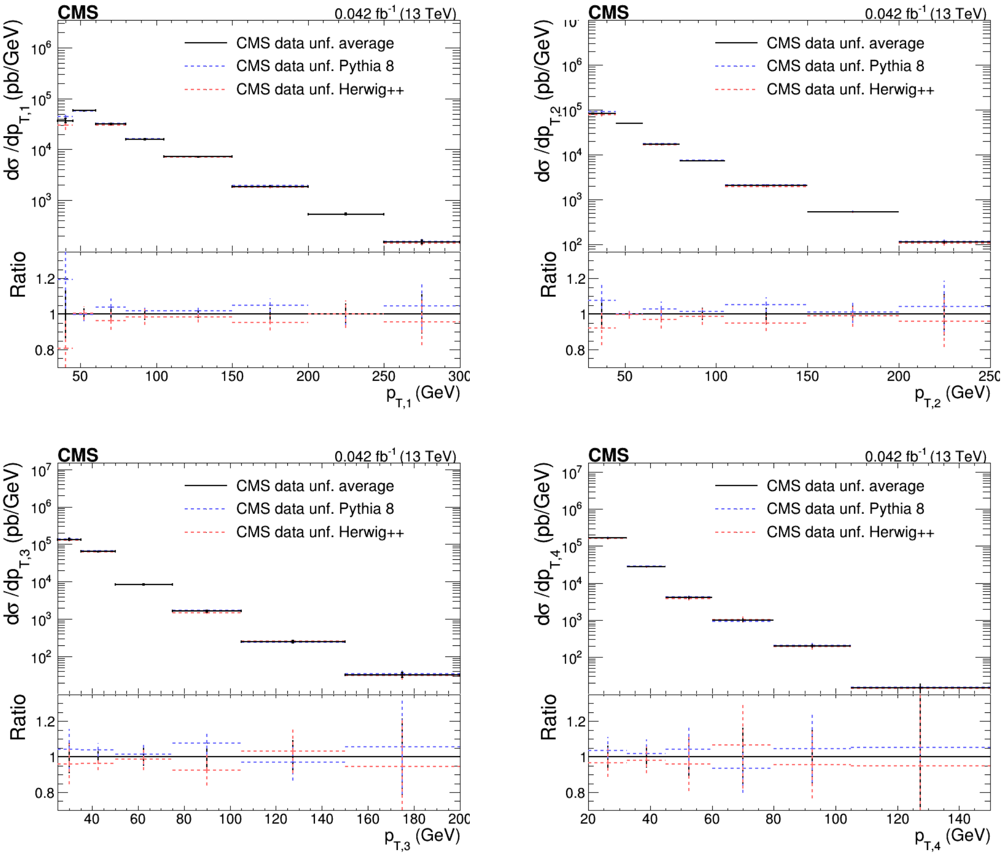


Figure 10.24: The p_{T} spectra unfolded with the PYTHIA 8 and HERWIG++ samples, along with the average unfolded p_{T} spectrum for the leading (top left), sub-leading (top right), third leading (bottom left) and fourth leading (bottom right) jets, are shown.

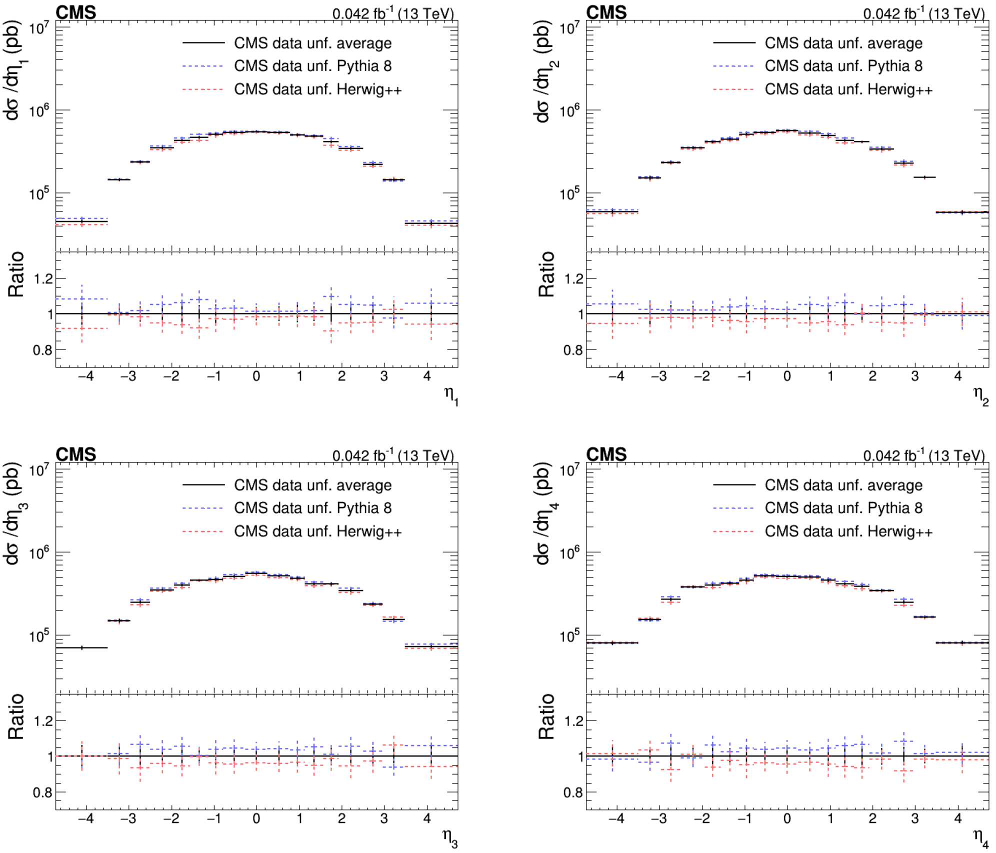


Figure 10.25: The η spectra unfolded with the PYTHIA 8 and HERWIG++ samples, along with the average unfolded η spectrum for the leading (top left), sub-leading (top right), third leading (bottom left) and fourth leading (bottom right) jets, are shown.

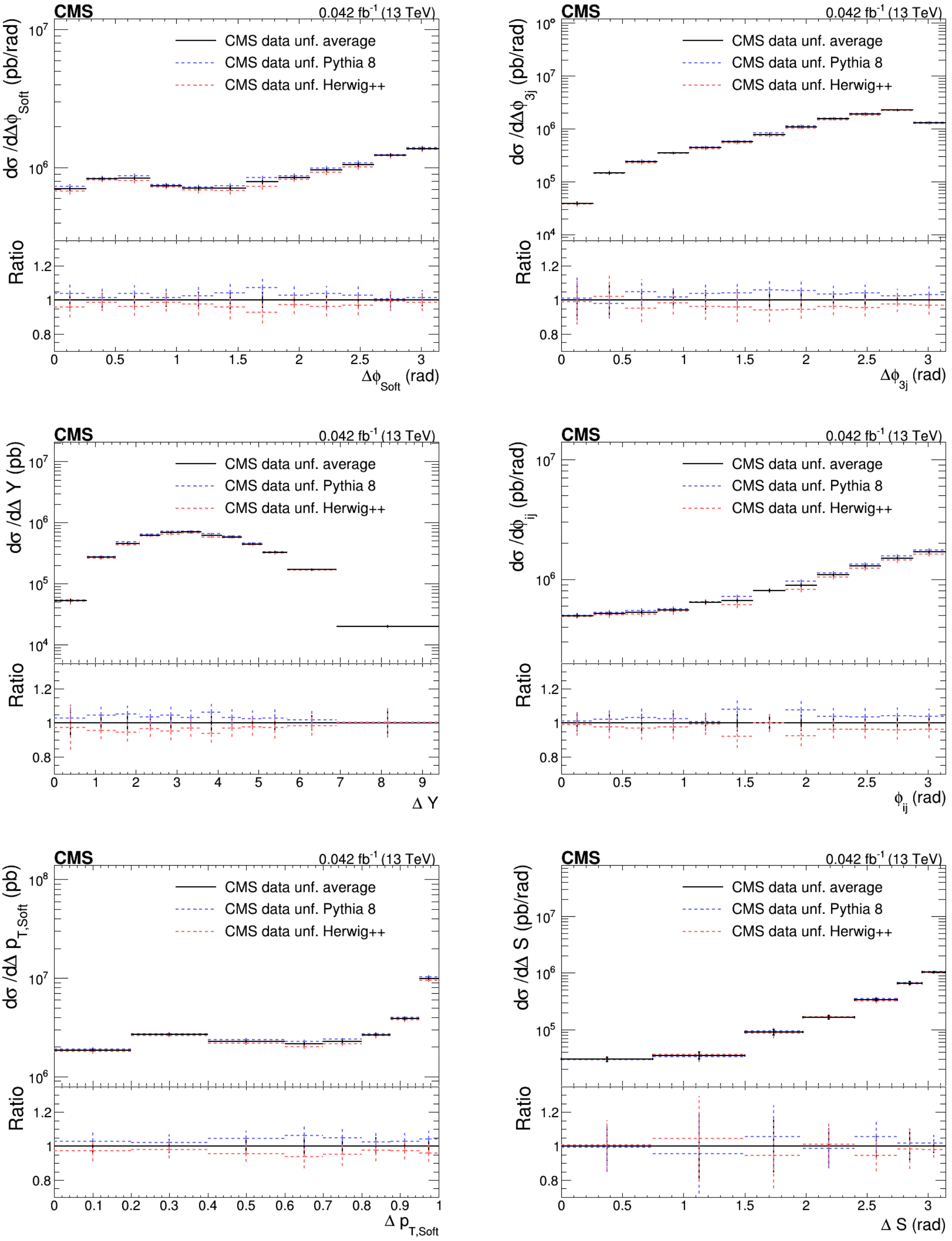


Figure 10.26: The DPS-sensitive observables unfolded with the PYTHIA 8 and HERWIG++ samples, along with the average unfolded distributions, are shown. The ΔS observable (bottom right) has been determined in *region II*, compared to *region I* for all other observables.

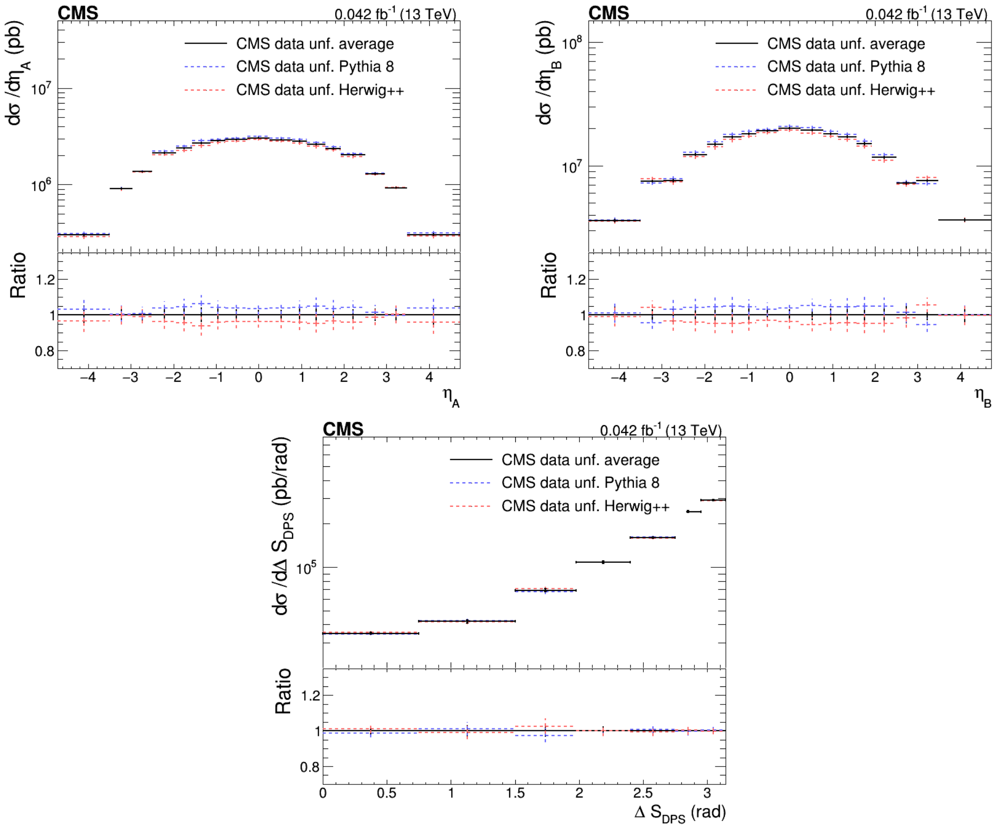


Figure 10.27: The η spectra of the inclusive single jet cross sections and the $\Delta\mathrm{S}_{\mathrm{DPS}}$ observable unfolded with the according PYTHIA 8 and HERWIG++ samples, alongside the average unfolded distribution, are shown.

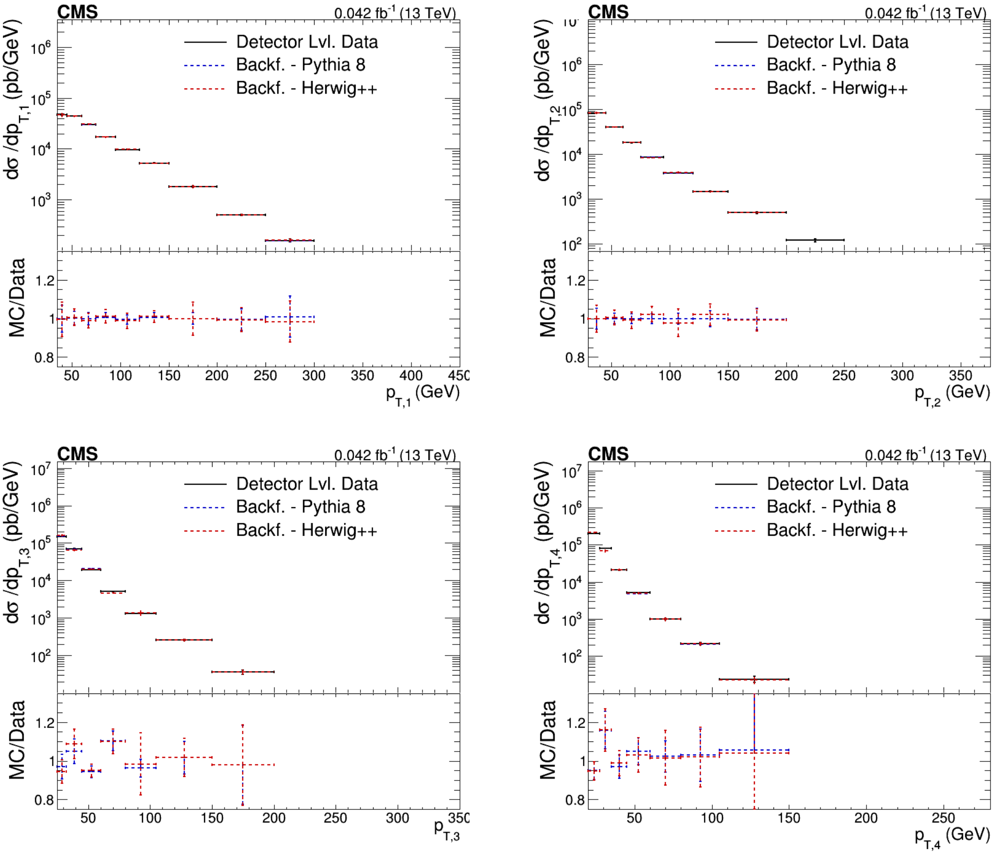


Figure 10.28: The backfolded p_{T} spectra are shown and compared to the detector level PYTHIA 8 and HERWIG++ distributions for the leading (top left), sub-leading (top right), third leading (bottom left) and fourth leading (bottom right) jets.

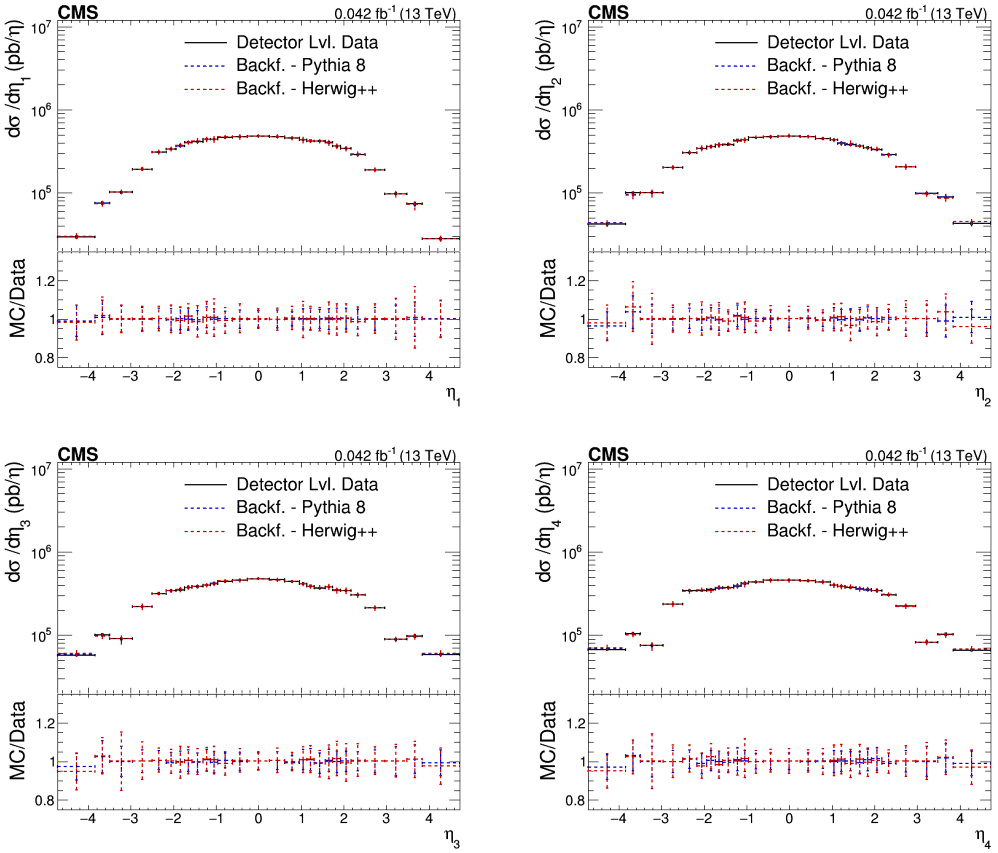


Figure 10.29: The backfolded η spectra are shown and compared to the detector level PYTHIA 8 and HERWIG++ distributions for the leading (top left), sub-leading (top right), third leading (bottom left) and fourth leading (bottom right) jets.

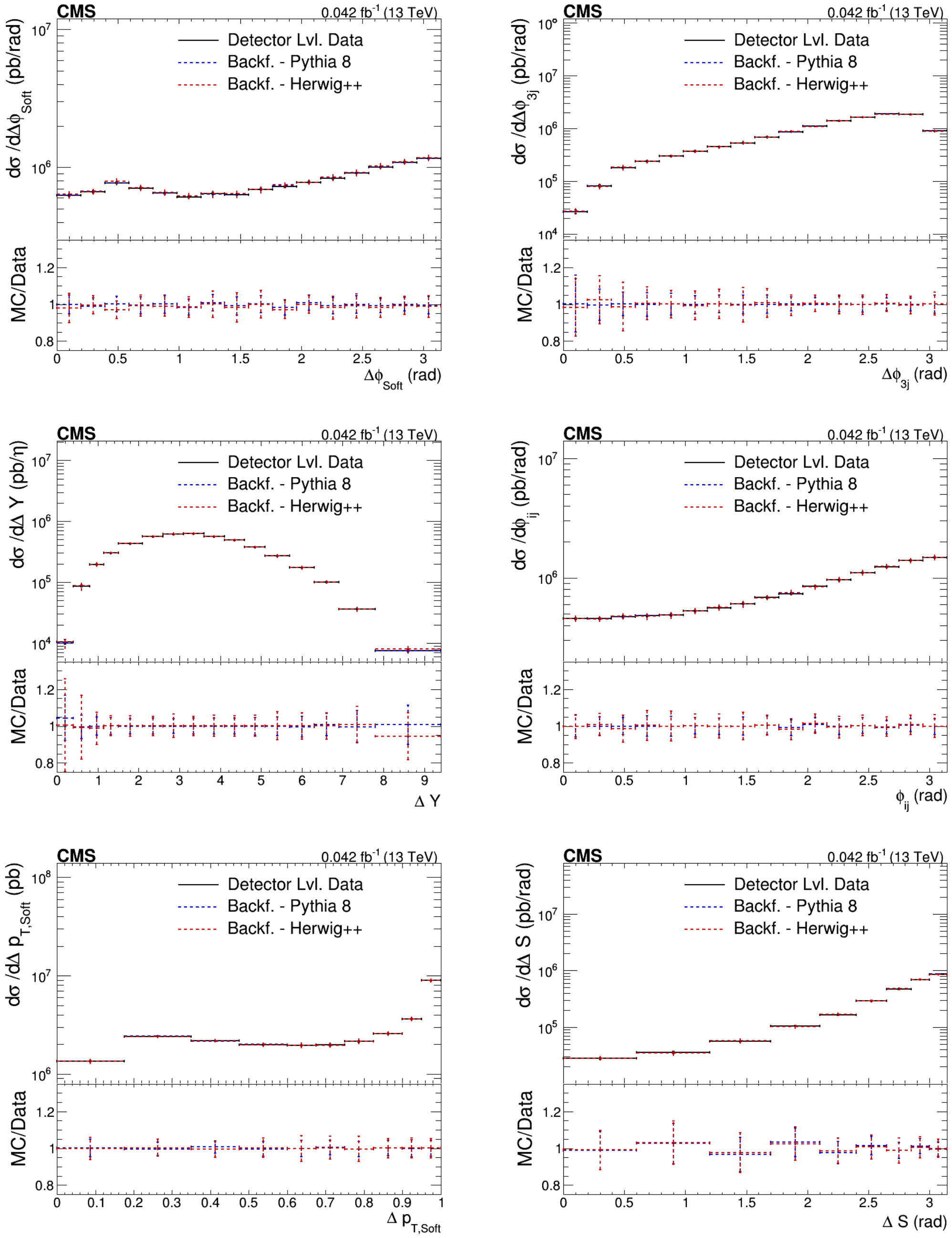


Figure 10.30: The backfolded DPS-sensitive observables are shown and compared to the detector level PYTHIA 8 and HERWIG++ distributions are shown here. The ΔS observable (bottom right) has been determined in *region II*, compared to *region I* for all other observables.

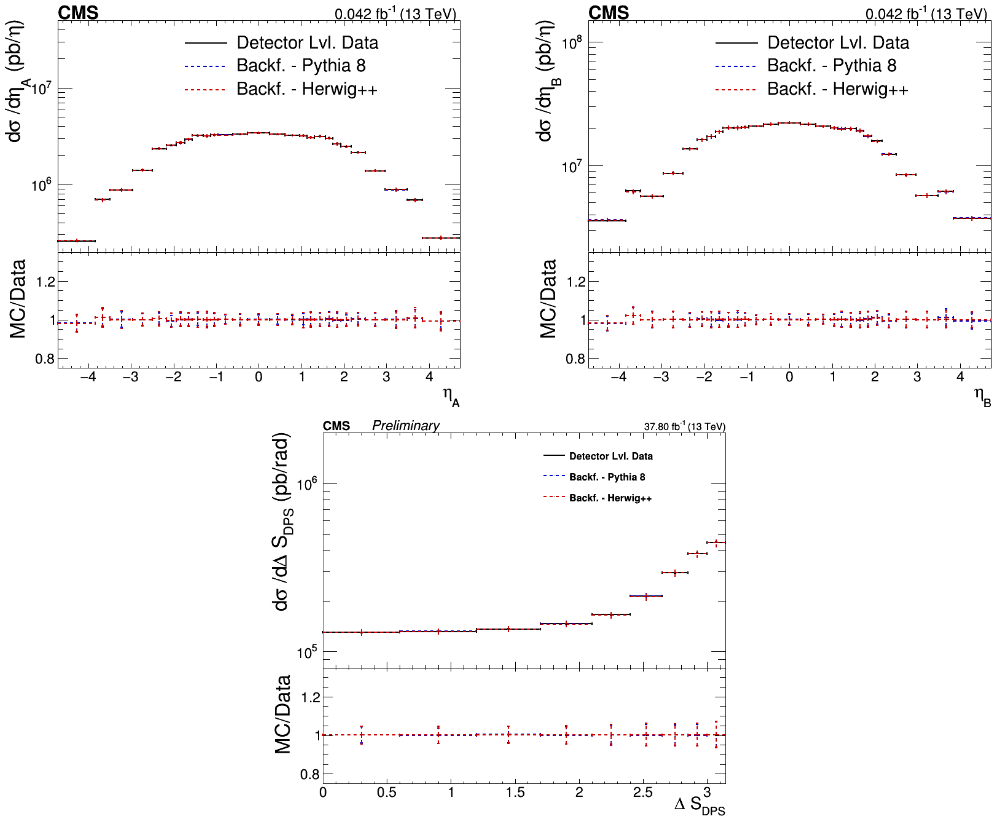


Figure 10.31: The backfolded η spectra for the inclusive single jet spectra and the $\Delta\mathrm{S}_{\mathrm{DPS}}$ observable are shown and are compared to the according detector level PYTHIA 8 and HERWIG++ distributions.

10.6 Cross Checks for the Unfolding Algorithm

The unfolding procedure is an indispensable part of the whole analysis. Any possible bias in the procedure could affect the unfolded results, therefore, a multitude of cross checks have been performed to ensure the robustness of the unfolding procedure.

A first cross check is a closure test of the unfolding procedure. The detector level PYTHIA 8 distribution is unfolded with the PYTHIA 8 sample itself. The unfolded PYTHIA 8 curve should agree quite well with the generator level distribution. If not, then something fundamental in the procedure has gone awry. The whole unfolding procedure is repeated as detailed in the previous section, where only the input distribution is changed to the detector level PYTHIA 8 distribution. The results are shown in Fig. 10.32, Fig. 10.33, Fig. 10.34 and Fig. 10.35. All distributions show perfect closure.

A second cross check is performed by unfolding the detector level PYTHIA 8 distribution with the HERWIG++ sample. One wishes for the unfolding procedure to be as model independent as possible. Agreement of the unfolded PYTHIA 8 curves with the generator level PYTHIA 8 distributions would indicate the robustness of the unfolding procedure. Substantial deviations would point towards a more significant model dependence, resulting in larger systematic errors due to the unfolding procedure. The systematic uncertainty will be revisited later on in Sec. 11.3. The results for all observables are shown in Fig. 10.36, Fig. 10.37, Fig. 10.38 and Fig. 10.39.

The p_T spectra of the four jets never deviate more than 10% from unity, except for the first bin of the second leading jet and the fourth bin of the last leading jet. The pseudorapidity spectra for the four leading jets show agreement within their statistical uncertainty in the central region. In the forward region, some outliers can be observed. Their deviations never exceed 20%, while the pseudorapidity spectra of the inclusive single jet spectra show perfect agreement throughout the whole pseudorapidity range. Of all other observables, the ΔS observable is the only one that does not show agreement within statistical uncertainty between the unfolded curve and the generator level distribution. The differences in the ΔS spectra are due to a broader migration matrix and due to the higher p_T cuts needed in order to obtain results with manageable errors. Overall one can conclude that the unfolding algorithm is quite robust. A model dependence is only observed in some bins of the p_T spectra and the ΔS observable and will be handled accordingly when calculating the systematic uncertainty resulting from the unfolding procedure.

A third and last cross check is performed by examining the reduced χ^2 of the smeared and unfolded space in the so-called bottom-line test. The χ^2 values are defined as

$$\chi^2_{\mathrm{sm}} = \left(\mathbf{y} - \mathbf{b} - \mathbf{A}\mathbf{x}'\right)^\mathrm{T} \mathbf{V_y}^{-1} \left(\mathbf{y} - \mathbf{b} - \mathbf{A}\mathbf{x}'\right) \tag{10.20}$$

$$\chi^2_{\mathrm{unf}} = \left(\mathbf{x} - \mathbf{x}'\right)^\mathrm{T} \mathbf{V_x}^{-1} \left(\mathbf{x} - \mathbf{x}'\right) \tag{10.21}$$

Apart from $\mathbf{x}'$, which is the generator level prediction by the MC model, all the definitions of the vectors and matrices are the same as before. The bottom-line test expresses that the unfolding procedure can only preserve the information about the model present in the data at best. In other words: the agreement between the models and the data can not become worse after the unfolding. Meaning that the inequality below should hold for all unfolded distributions.

$$\chi^2_{\mathrm{sm}}/\mathrm{NDF}_{\mathrm{sm}} \geq \chi^2_{\mathrm{unf}}/\mathrm{NDF}_{\mathrm{unf}} \tag{10.22}$$

Where $\mathrm{NDF}_{\mathrm{sm}}$ and $\mathrm{NDF}_{\mathrm{unf}}$ are the number of degrees of freedom in the smeared and unfolded space, respectively. They are in turn determined by the rank of the matrices $\mathbf{A}\mathbf{V_y}\mathbf{V_y}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}\mathbf{V_x}\mathbf{V_x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$, respectively.

All reduced χ^2 values for all the observables are given in Tab. 10.1. The NDF is the same for the χ^2_{backf} and the χ^2_{sm} sine they are both at detector level. From the calculation of the rank of the matrices $\mathbf{A}\mathbf{V_y}\mathbf{V_y}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}\mathbf{V_x}\mathbf{V_x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$, it appears that the rank of both matrices is the same and equal to the number of generator bins. One can also conclude that the bottom-line test is always fulfilled.

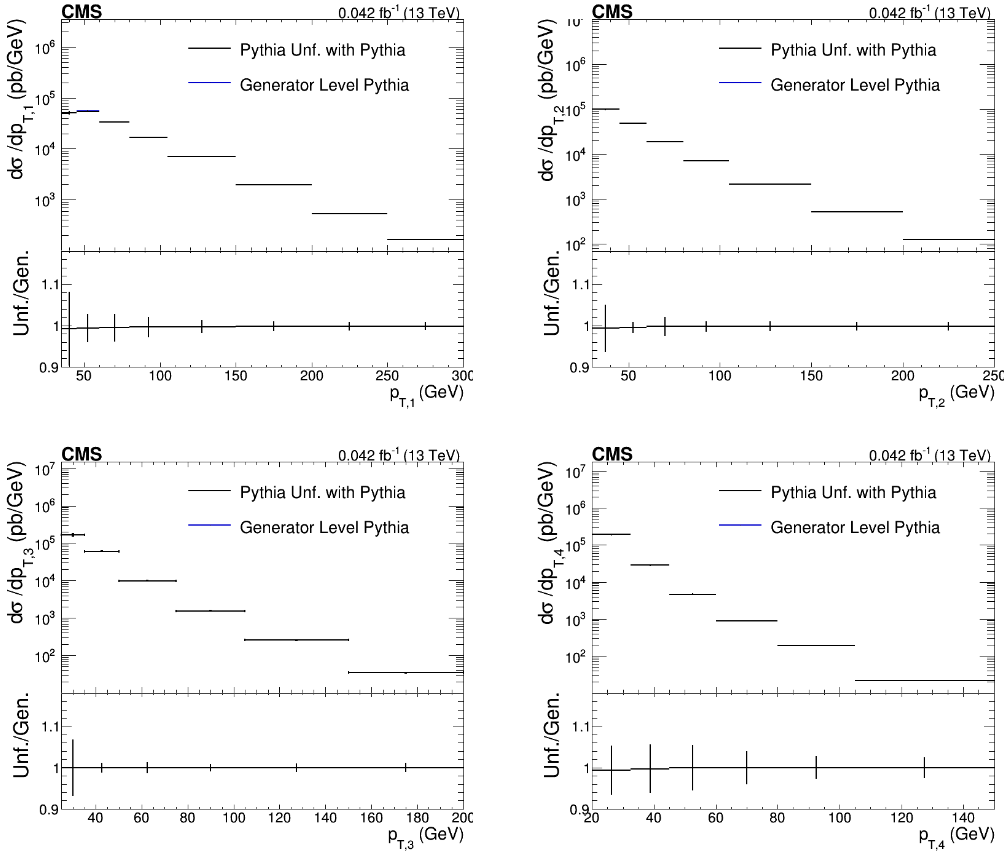


Figure 10.32: The closure test for the unfolding procedure for the p_{T} distributions for the leading (top left), sub-leading (top right), third leading (bottom left) and fourth leading (bottom right) jets is shown.

Table 10.1: The values of the NDF in the smeared and unfolded space, together with the minimal reduced χ^2 values for the backfolding and with the associated reduced χ^2 values of the smeared and unfolded space are all shown in the table below. The condition number, whether or not the additional area constraint is used and the regularization method are also given for all unfolded distributions.

Observable	$\mathrm{NDF}_{\mathrm{sm}}$	Min.Red. χ^2_{backf}	cond($\mathbf{A}$)	Area Const.	Reg. Cond.	Red. χ^2_{sm}	$\mathrm{NDF}_{\mathrm{unf}}$	Red. χ^2_{unf}
Unfolded with the PYTHIA 8 sample								
$p_{\mathrm{T,1st}}$	8	0.0166	5.76	Yes	Size	16.5102	8	16.2167
$p_{\mathrm{T,2nd}}$	7	0.0007	13.59	No	Der.	20.2665	7	8.9505
$p_{\mathrm{T,3rd}}$	6	1.5426	23.48	Yes	Size	96.8042	6	60.0743
$p_{\mathrm{T,4th}}$	6	0.9405	81.85	Yes	Size	120.2859	6	61.2052
η_1	17	0.0039	8.40	Yes	Size	0.4083	17	0.3744
η_2	17	0.0404	5.98	Yes	Size	3.1776	17	2.8358
η_3	17	0.0305	5.72	Yes	Size	3.5137	17	3.0968
η_4	17	0.0426	6.33	Yes	Size	3.4488	17	2.9382
$\Delta\phi_{\mathrm{soft}}$	12	0.0352	2.81	Yes	Size	0.5395	12	0.2856
$\Delta\phi_{\mathrm{3j}}^{\mathrm{min}}$	12	0.0008	71.00	Yes	Size	0.3722	12	0.3669
$\Delta\mathrm{Y}$	12	0.0142	10.79	Yes	Size	1.0708	12	1.0321
ϕ_{ij}	12	0.0134	4.55	Yes	Size	3.2838	12	3.1658
$\Delta p_{\mathrm{T,soft}}$	8	0.0105	4.28	Yes	Size	0.9724	8	0.8460
$\Delta\mathrm{S}$	7	0.1079	32.12	Yes	Size	1.3929	7	0.7184
$\Delta\mathrm{S}_{\mathrm{DPS}}$	7	0.0234	8.46	Yes	Size	34.3008	7	34.1627
η_{A}	17	0.0284	6.23	Yes	Size	1.6404	17	1.1321
η_{B}	17	0.0511	4.44	Yes	Size	154.7417	17	148.6421
Unfolded with the HERWIG++ sample								
$p_{\mathrm{T,1st}}$	8	0.0525	5.64	Yes	Size	85.5236	8	82.3845
$p_{\mathrm{T,2nd}}$	7	0.3148	14.38	Yes	Size	69.8411	7	29.2084
$p_{\mathrm{T,3rd}}$	6	1.2485	21.71	Yes	Size	153.0023	6	60.7456
$p_{\mathrm{T,4th}}$	6	0.6857	85.37	Yes	Size	106.2534	6	46.6522
η_1	17	0.0135	7.98	Yes	Size	1.5272	17	1.2242
η_2	17	0.0654	5.83	Yes	Size	1.7157	17	1.0028
η_3	17	0.0401	5.91	Yes	Size	2.2637	17	0.9845
η_4	17	0.0595	6.17	Yes	Size	2.5032	17	1.2591
$\Delta\phi_{\mathrm{soft}}$	12	0.0363	2.77	Yes	Size	1.2991	12	0.9432
$\Delta\phi_{\mathrm{3j}}^{\mathrm{min}}$	12	0.0106	70.82	Yes	Size	1.2024	12	0.9957
$\Delta\mathrm{Y}$	12	0.0173	10.42	Yes	Size	1.8080	12	1.4310
ϕ_{ij}	12	0.0209	4.45	Yes	Size	0.7883	12	0.4216
$\Delta p_{\mathrm{T,soft}}$	8	0.0041	4.35	Yes	Size	1.0853	8	0.9481
$\Delta\mathrm{S}$	7	0.0409	29.94	Yes	Size	1.1566	7	0.6435
$\Delta\mathrm{S}_{\mathrm{DPS}}$	7	0.0245	8.31	Yes	Size	24.5340	7	24.3878
η_{A}	17	0.0188	6.07	Yes	Size	2.3252	17	1.8330
η_{B}	17	0.0312	4.11	Yes	Size	112.4995	17	108.6421

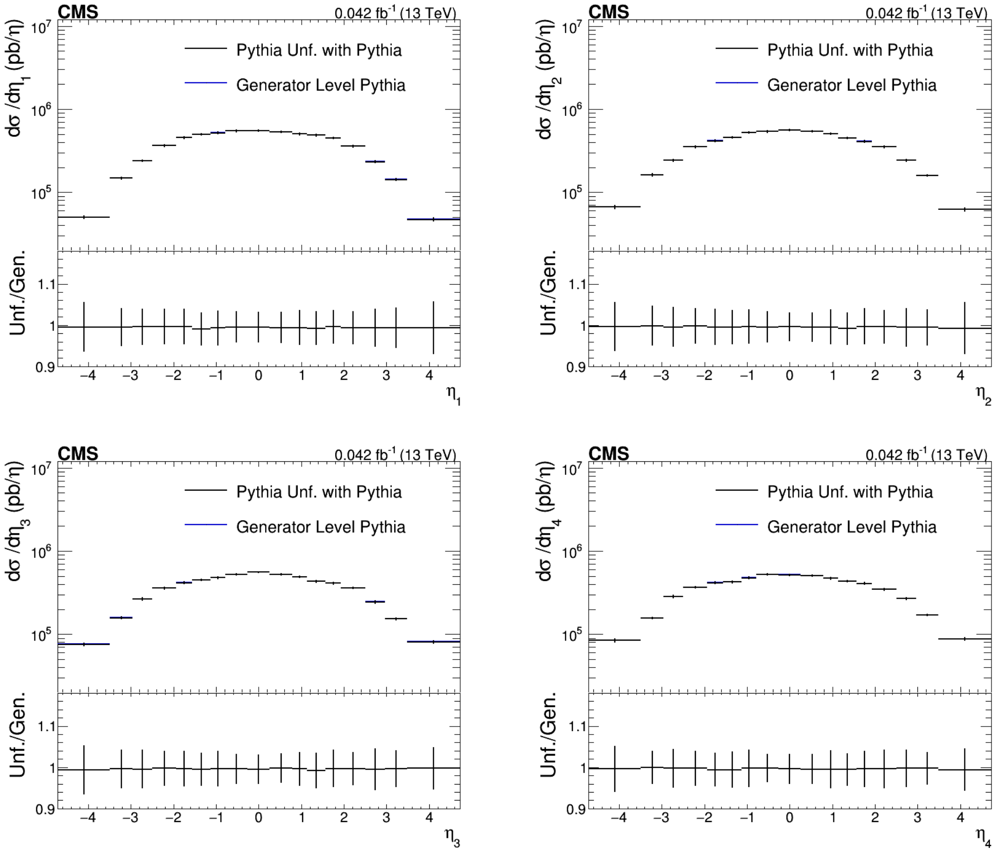


Figure 10.33: The closure test for the unfolding procedure for the η distributions for the leading (top left), sub-leading (top right), third leading (bottom left) and fourth leading (bottom right) jets is shown.

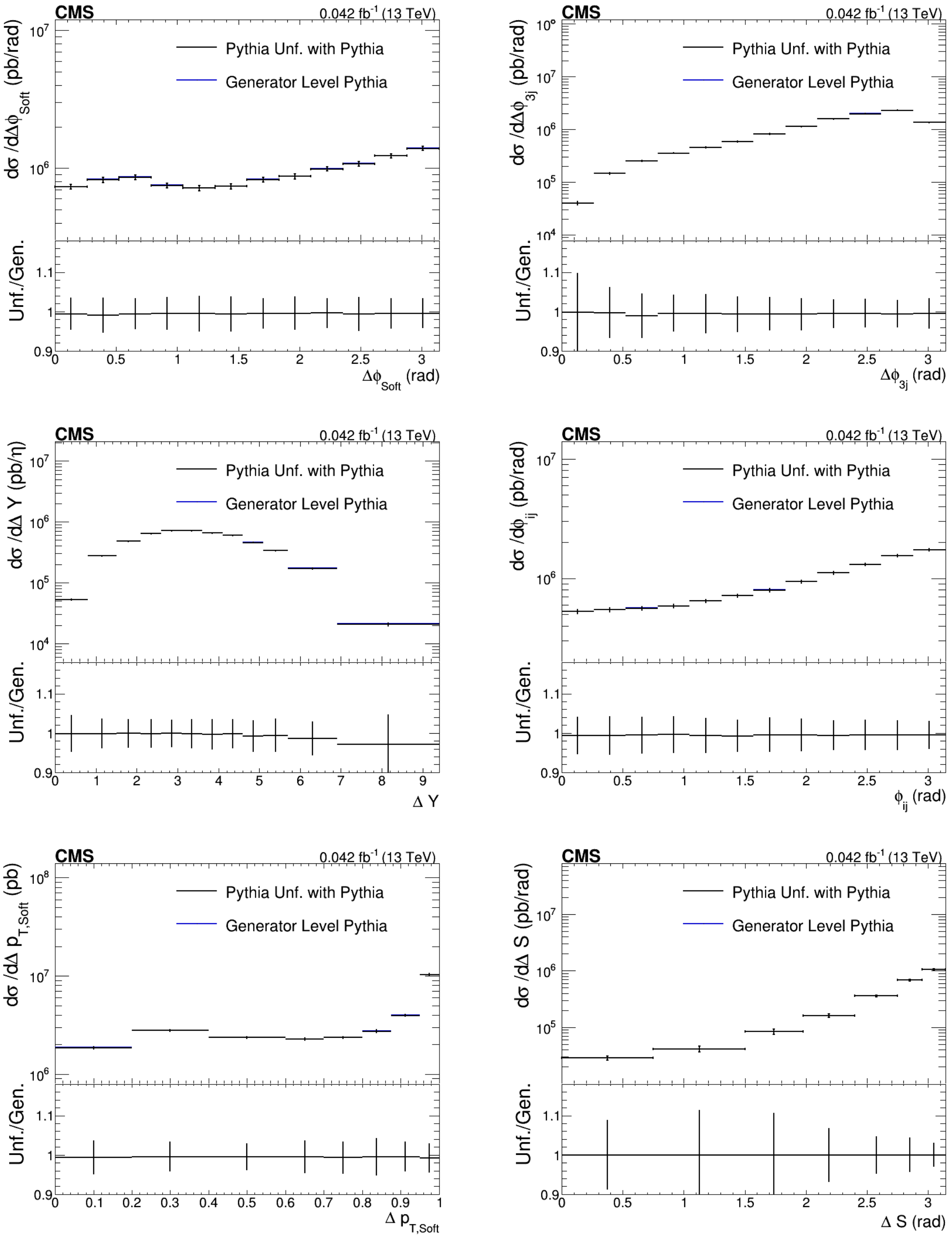


Figure 10.34: The closure test for the unfolding procedure for the DPS-sensitive observables is shown. The ΔS observable (bottom right) has been determined in *region II*, compared to *region I* for all other observables.

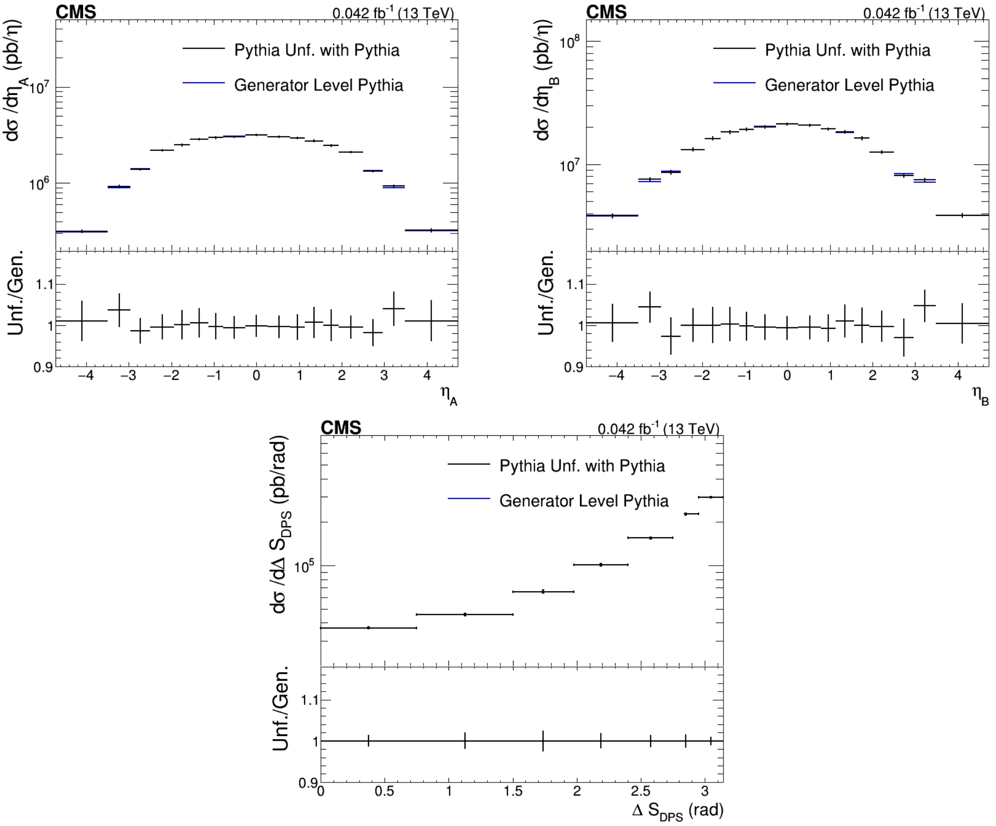


Figure 10.35: The closure test for the unfolding procedure for the η spectra for the inclusive single jet spectra and the ΔS_{DPS} observable is shown.

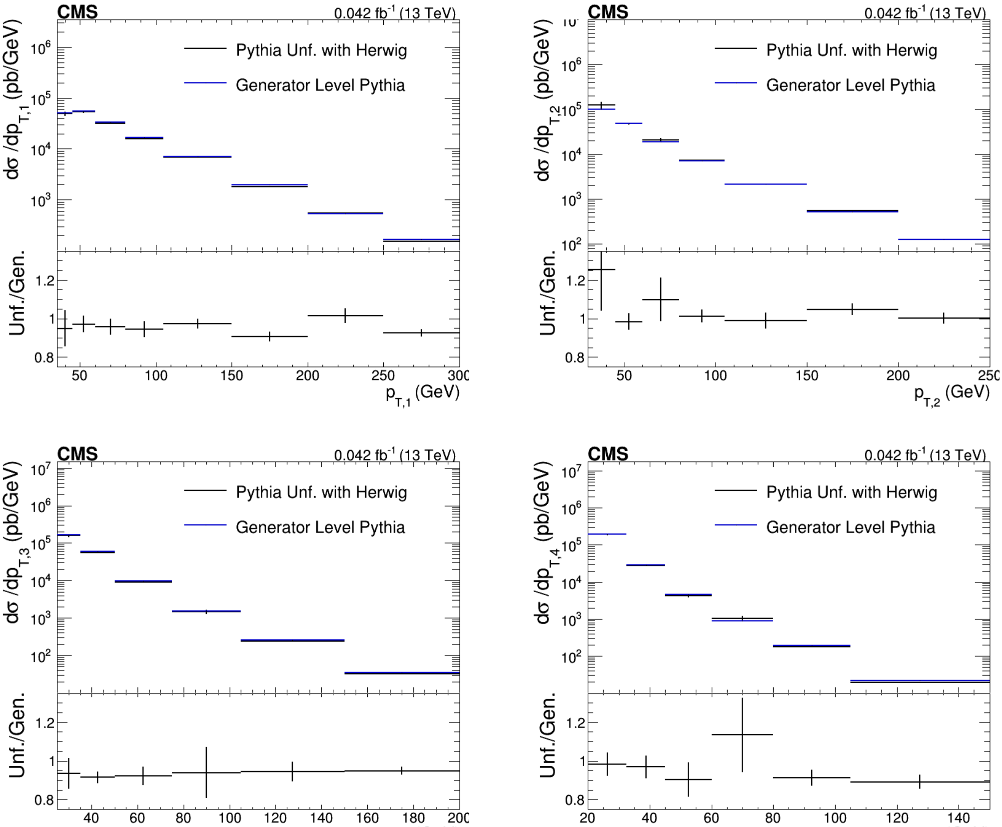


Figure 10.36: The model dependence cross check of the unfolding procedure of the p_{T} spectra for the leading (top left), sub-leading (top right), third leading (bottom left) and fourth leading (bottom right) jets is shown.

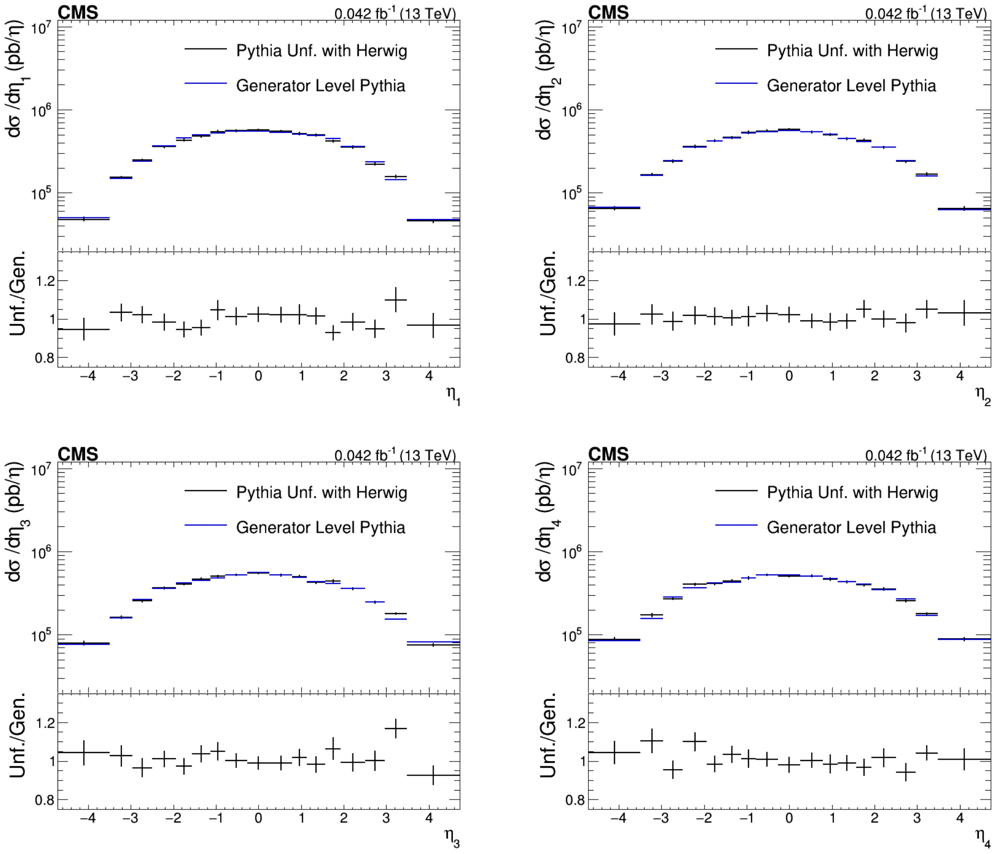


Figure 10.37: The model dependence cross check of the unfolding procedure of the η spectra for the leading (top left), sub-leading (top right), third leading (bottom left) and fourth leading (bottom right) jets is shown.

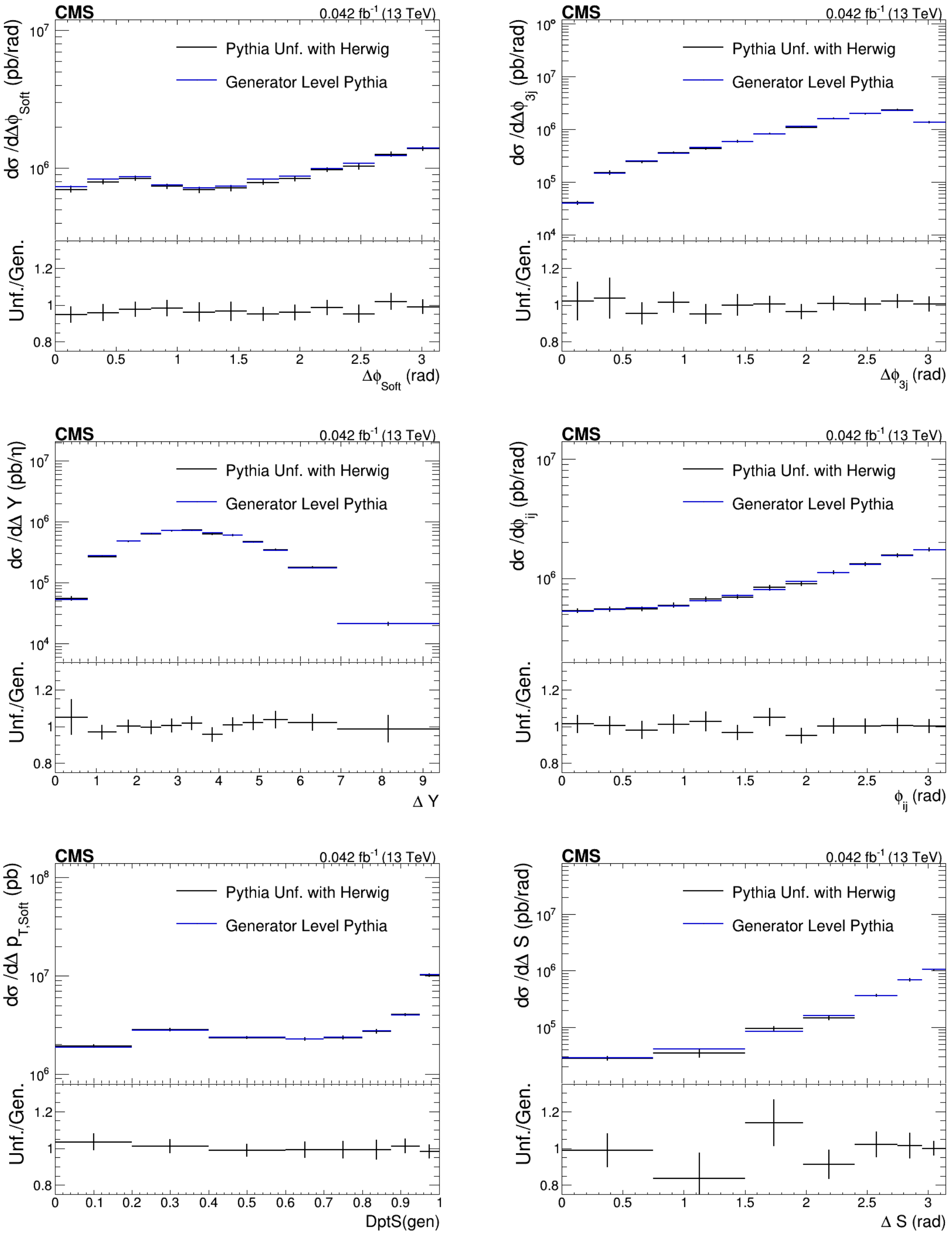


Figure 10.38: The model dependence cross check of the unfolding procedure of the DPS-sensitive observables is shown. The ΔS observable (bottom right) has been determined in *region II*, compared to *region I* for all other observables.

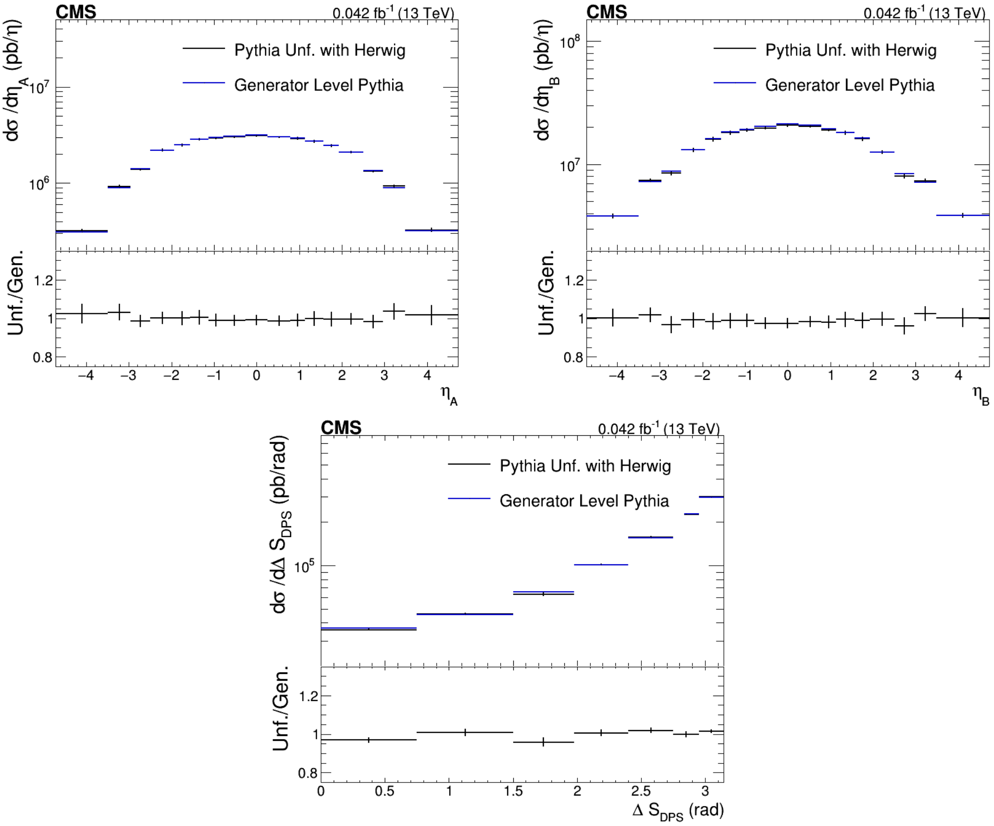


Figure 10.39: The model dependence cross check of the unfolding procedure of the η spectra for the inclusive single jet spectra and the $\Delta\mathrm{S}_{\mathrm{DPS}}$ observable is shown.

Chapter 11

Systematic Uncertainties

Any analysis is subjected to additional uncertainties besides the statistical uncertainty. The systematic uncertainties all result from certain operations performed in the analysis. All sources of systematic uncertainties will be discussed in what follows, except the systematic uncertainty due to the measurement of the luminosity. It is a fixed value for each data taking period and is determined by the CMS experiment. For all of the data taken in 2016, the uncertainty on the luminosity is 2.5% [182]. After discussing all other sources of systematic uncertainties, the total systematic uncertainty for each observable will be determined and discussed.

11.1 Jet Energy Scale Uncertainty

The first systematic uncertainty is the Jet Energy Scale (JES) uncertainty. The reconstructed jets undergo a series of sequential corrections in the form of a scaling of their four-momentum vector in order to obtain the calibrated jets, schematically shown in Fig. 6.1. The corrections carry an uncertainty themselves. Additionally, a non-closure of the L2L3 residual corrections was found due to the limited available statistics. Therefore, it was opted to apply the L2L3 residual corrections and to treat the possible non-closure as an additional uncertainty. If a jet with a given p_{T} and η exhibits non-closure, the total error on the JEC becomes

$$s_{\mathrm{JEC}}^{\mathrm{full}}(p_{\mathrm{T}},\eta) = \sqrt{s_{\mathrm{JEC}}^{2}(p_{\mathrm{T}},\eta) + s_{\text{non-cl}}^{2}(p_{\mathrm{T}},\eta)} \tag{11.1}$$

Where s_{JEC} is the total error of all the JEC factors combined. The magnitude of the non-closure $s_{\text{non-cl}}$ is equal to the deviation from unity in the closure plot for the MPF method in Fig. 6.9. In order to determine the JEC uncertainty, the JEC factors are varied up and down by their respective uncertainty. Afterwards, the whole analysis is repeated. Next, the varied distributions are determined and compared to the nominal distributions. Lastly, the difference between the varied and nominal distributions is taken as the JES uncertainty. As the upward and downward variation of the JEC factors differs, the JES uncertainty is an asymmetric uncertainty.

The resulting JES uncertainties for all the observables are shown in Fig. 11.6, Fig. 11.7, Fig. 11.8 and Fig. 11.9. The JES uncertainty for the upward (downward) variation of the JEC factors is labeled as the *upper* (*lower*) JES uncertainty. Looking at the JES uncertainty for the p_{T} spectra of the jets, one observes large values at low p_{T}, up to 39% for the leading jet, while the JES uncertainty decreases to values as low as 8% for high values of the p_{T}. The JES uncertainty for the pseudorapidity spectra and for the other observables do not show as much variation compared to the p_{T} spectra. Generally, the upper uncertainty is larger as the lower uncertainty. Both range between 2% and 39% for all observables. As stricter p_{T} cuts are in place for the variable ΔS, the JES uncertainty is lower in comparison to all the other variables.

In order to understand the effect of the non-closure, the same uncertainties have been calculated without taking the non-closure into account. In Fig. 11.1 the relative JES uncertainties of the p_{T} spectrum of the two leading jets are shown with and without the non-closure quadratically added to the uncertainty of the JES factors. The effect of the non-closure compared to the size of the systematical uncertainties is small throughout the p_{T} spectra. The non-closure is not the main contributing factor to the JES uncertainty.

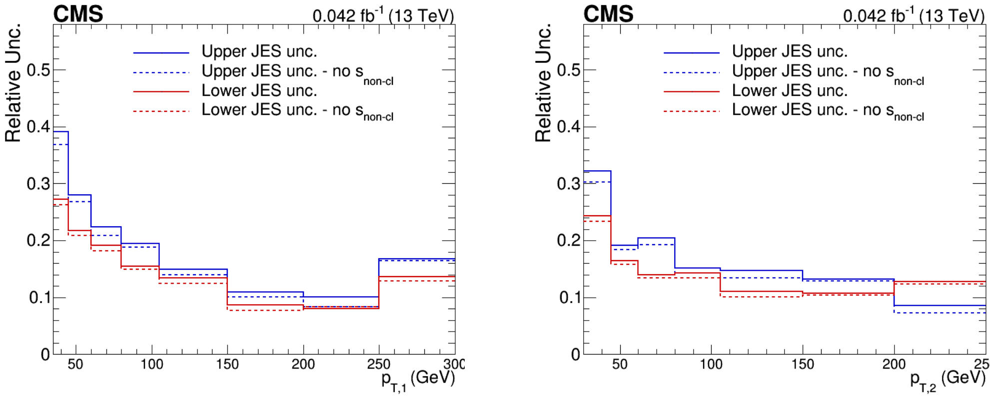


Figure 11.1: The relative JES uncertainty for the p_{T} spectrum of the two leading jets with and without the non-closure taken into account. The uncertainties have been obtained by varying the p_{T} of the jets up and down by the total error of the JEC factors.

11.2 Jet Energy Resolution Uncertainty

The Jet Energy Resolution (JER) from MC simulations differs from the one estimated for data, which would lead to a wrong estimation of the bin-to-bin migrations. An additional smearing of the p_{T} of the jets at detector level is applied to the MC samples, used in the unfolding. The two smearing methods used to perform the additional smearing are detailed below.

- **The scaling method**: the method only works properly when a jet on detector and generator is matched in p_{T} with the criteria given below.

$$\left((\phi^{\text{gen}} - \phi^{\text{det}})^2 + (\eta^{\text{gen}} - \eta^{\text{det}})^2\right)^{\frac{1}{2}} < \frac{R_{\text{cone}}}{2} \tag{11.2}$$

$$\text{and} \quad \left|\frac{p_{\text{T}}^{\text{gen}} - p_{\text{T}}^{\text{det}}}{p_{\text{T}}^{\text{gen}}}\right| < 3 \cdot \sigma_{\text{JER}} \tag{11.3}$$

Where R_{cone} is the cone radius of the jets, which is equal to 0.4, and σ_{JER} is the relative p_{T} resolution as measured in the simulation. If the criteria are met then jets are smeared with the smearing factor c_{JER}, defined as

$$c_{\text{JER}} = 1 + (s_{\text{JER}} - 1)\frac{p_{\text{T}}^{\text{det}} - p_{\text{T}}^{\text{gen}}}{p_{\text{T}}^{\text{det}}} \tag{11.4}$$

The factor s_{JER} is the data-to-simulation scale factor, provided by the CMS Collaboration [183] and obtained from the global tag which is specified in Chap. 7.

- **The stochastic method**: the scale factor is determined by using a random number sampled from a normal distribution N. The smearing factor is given by

$$c_{\text{JER}} = 1 + N(0, \sigma_{\text{JER}})\sqrt{\max((s_{\text{JER}} - 1), 0)} \tag{11.5}$$

If no match is found for the generator level jet on detector level, then the stochastic method is used. Both methods combined form the hybrid smearing method and are truncated at zero.

The uncertainty of the data-to-simulation scale factor s_{JER} gives rise to the JER uncertainty. In order to obtain the JER uncertainty, the data-to-simulation scale factor s_{JER} is varied up and down with its respective uncertainty, resulting in migration matrices that differ from their nominal counterparts. The newly obtained matrices are used to unfold the distributions, which are then compared to the nominal distributions for all observables. The JER uncertainty is taken as the maximum deviation of the varied distributions compared to the nominal distribution.

The resulting JER uncertainties are shown in Fig. 11.6, Fig. 11.7, Fig. 11.8 and Fig. 11.9. The JES uncertainty never exceeds 9% for all observables, except for the p_{T} spectrum of the leading jet, where it reaches a maximum of 26%.

11.3 Model Uncertainty

The unfolding procedure is model dependent since the migration matrix is determined by the MC models. As discussed in the previous chapter, the data is unfolded using two independent models. The average of both unfolded curves is taken as the final distribution in order to reduce the bias of the unfolded data towards one model or another. The systematic error due to the model dependence of the unfolding procedure is taken as half of the difference between the two unfolded distributions.

In Fig. 11.6, Fig. 11.7, Fig. 11.8 and Fig. 11.9, the relative model uncertainties for all observables are shown. The model uncertainty varies between 1% and 16%, depending on the variables.

11.4 Trigger Uncertainty

The trigger efficiency curves have been determined in Sec. 9.1 since the event weight needs to be corrected for possible trigger inefficiencies as a function of the p_{T} of the jets. The curves have been parameterized and fitted by an error function given in eq. (9.2). However, the fitting parameters have a non-zero uncertainty. The trigger uncertainty is estimated by varying the parameters up and down with their respective uncertainty, whereafter, the whole analysis is repeated with the varied efficiency curves. Lastly, the varied unfolded distributions are compared to their nominal distributions.

The nominal and varied efficiency curves for the `AK4PFJetXX` triggers are shown in Fig. 11.2. Only small deviations in the slope of the curves are observed compared to the nominal efficiency function. In Sec. 7, the trigger configuration was laid out. All triggers, except the `AK4PFJet30` trigger, are used in the regions where they have become fully efficient. The uncertainty on the fitting parameters are the smallest for the `AK4PFJet30` trigger. Therefore, the effect of the varied efficiency curves on the distributions of the observables will be minimal. The systematic uncertainties due to the trigger efficiency are shown in Fig. 11.3 for the p_{T} spectrum of the two leading jets. The uncertainties are of sub-percent level and are negligible compared to the other uncertainties. In Fig. 11.6, Fig. 11.7, Fig. 11.8 and Fig. 11.9, the trigger uncertainties are omitted for clarity's sake, however, they are included in the calculation of the total uncertainty.

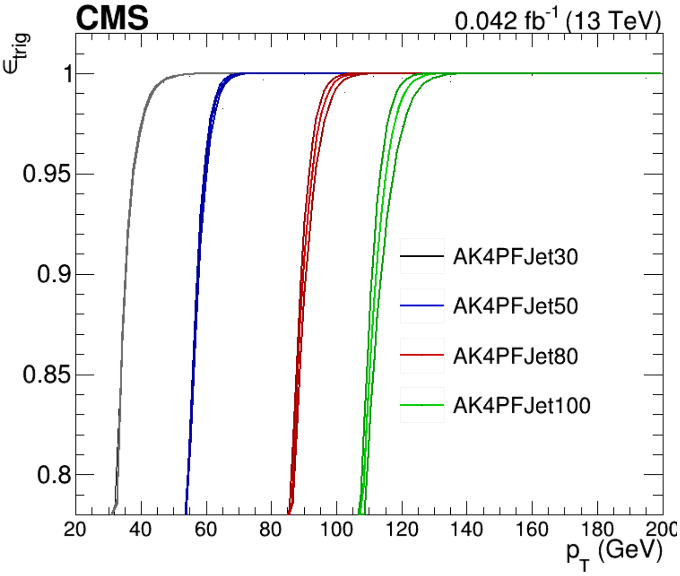


Figure 11.2: The trigger efficiency curves together with the curves for which the fitting parameters have been varied up and down with their respective uncertainty. The nominal efficiency curves are the middle ones for each of the triggers.

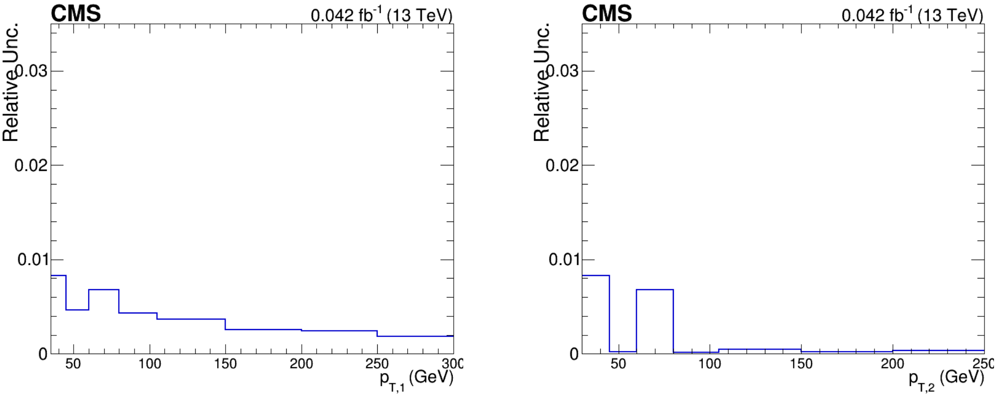


Figure 11.3: The trigger uncertainty for the p_T spectrum of the two leading jets.

11.5 Vertex Selection Uncertainty

When the offline selection is performed on the data, the presence of exactly one proper vertex is required. If an event contains zero or more than one proper vertex, it is discarded. The offline selection carries an uncertainty with it because the probability that two vertices are reconstructed as one is non-zero. The CMS detector has a high resolution and can differentiate between two distinct vertices as long as they are further apart than the minimal vertex separation, defined below.

$$\Delta z_{\mathrm{min}} = \min_{i \neq j} \{|z_i - z_j|\} \tag{11.6}$$

Where z_i and z_j are the z-coordinates of the distinct vertices in the same event. In Fig. 11.4, the Δz distribution determined from the data at detector level is shown. The distribution of the vertex separation confirms an older measurement, where a Δz_{min} of 0.06 cm [184] was measured. To be more conservative, however, a value of 0.12 cm is used since the curve exhibits a turning point at that value.

The effect of merging vertices in events needs to be mimicked in order to estimate the vertex selection uncertainty. First, the vertex merging rate is determined. The z-coordinate distribution of the vertices, depicted in Fig. 11.5, is fitted with a Gaussian function of which the mean and the variance are equal to zero and 3.906 cm, respectively. The Gaussian distribution of the distance between two vertices along the z-axis is determined by combining the initial distribution with itself. The mean and variance of the combined Gaussian distribution is equal to zero and $\sqrt{2} \cdot 3.906 = 5.52$ cm, respectively. The probability of encountering an event where two vertices have been merged, i.e. the vertex merging rate, can be expressed as

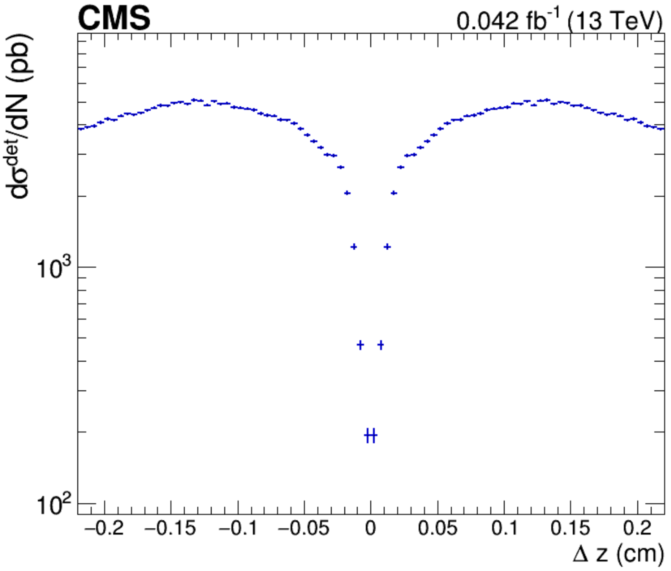


Figure 11.4: The Δz distribution determined from data at detector level. Whenever there are two or more reconstructed vertices present in an event, the difference between their z-coordinates is taken, to obtain the separation between the distinct vertices.

$$\begin{aligned} \mathrm{P}(|\mathrm{vtx}_1 - \mathrm{vtx}_2| \leq 0.12 \text{ cm}) &= \frac{0.23}{0.71} \frac{1}{\sqrt{2\pi} \cdot 5.52} \int_{-0.06}^{+0.06} \exp\left(\frac{-z^2}{2 \cdot 5.52^2}\right) dz \\ &= 0.0028 \end{aligned} \tag{11.7}$$

The first fraction expresses the conditional probability of the presence of a second vertex whenever a first vertex is present in the event. It is obtained from the vertex multiplicity distribution, shown in Fig. 7.1. In 71% and 23% of all events, one and two vertices have been reconstructed, respectively. Next, the whole analysis is repeated with an altered offline vertex selection criterion. Two proper vertices are required to be present in an event instead of one. All altered distributions are unfolded in exactly the same manner as before. Finally, combining the vertex merging rate with the one and two vertex distributions, allows for the estimation of the effect of the vertex merging by considering the expression below.

$$\frac{\sigma(1 \text{ vtx}) - 0.0028 \cdot \sigma(2 \text{ vtx})}{\sigma(1 \text{ vtx})} \tag{11.8}$$

Where the two vertex cross sections have been scaled to the one vertex cross section before the multiplication with the vertex merging rate. The vertex selection uncertainty is the absolute value of the deviation of the ratio to one.

The resulting uncertainties are shown in Fig. 11.6, Fig. 11.7, Fig. 11.8 and Fig. 11.9 for all observables. The vertex selection uncertainty never exceeds 1% and is negligible compared to other uncertainties.

In the calculation of the vertex merging rate, the x- and y-coordinates have been omitted for two reasons. First, they are much smaller as the z-coordinate. The selection criterion on the radial coordinate $\rho = \sqrt{x^2 + y^2}$ demands that $\rho \leq 2$ cm. Second, if the x- and y-coordinates would have been taken into account, the integral boundaries in eq. (11.7) would be closer to zero, effectively reducing the merging probability.

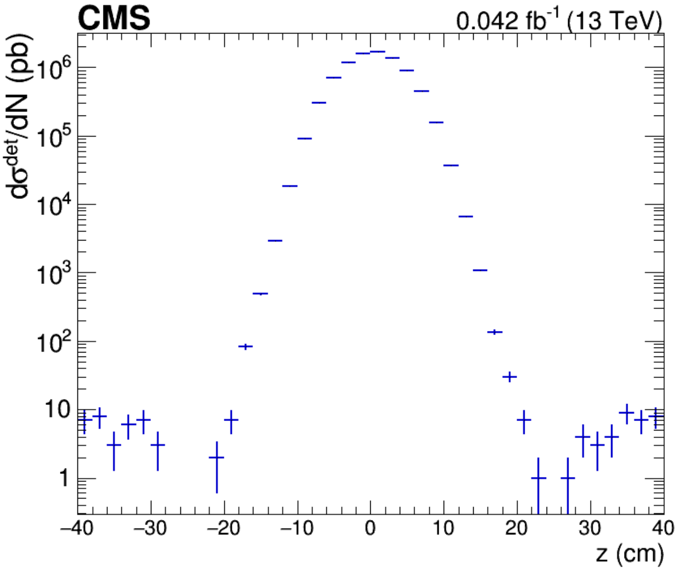


Figure 11.5: The vertex z-coordinate distribution used in the calculation of the vertex selection uncertainty.

11.6 Additional Uncertainty on the 4-Jet Efficiency

In Chap. 8, the 4-jet efficiency ϵ_{4j} was introduced as a ratio on detector level in order to reformulate the DPS pocket formula in terms of the inclusive single jet processes A and B. Possible differences between detector and generator level could affect the 4-jet efficiency. The PYTHIA 8 and HERWIG++ samples using the `CUETP8M1` and `CUETHS1` tunes, respectively, are used in this regard. The inclusive single jet cross sections are extracted from the samples. Afterwards, a two-dimensional reweighing is performed as a function of the p_T of the leading jet and of the jet multiplicity in order to let the MC samples be more representative of the data. Both the p_T spectrum of the leading jet and jet multiplicity play a crucial role in the whole mixing procedure. On the one hand, if the average p_T is larger in the MC sample compared to data, then more jets will surpass the p_T cuts, effectively increasing the 4-jet efficiency. On the other hand, a higher jet multiplicity will result in a rise of the 4-jet efficiency as well since events with multiple jets passing the p_T cuts will be prevalent. The generator and detector level 4-jet efficiencies are given in Tab. 11.1 for the reweighted PYTHIA 8 sample.

Table 11.1: The generator and detector level values of the 4-jet efficiencies determined from the PYTHIA 8 and HERWIG++ samples using the `CUETP8M1` and `CUETHS1` tunes, respectively. Before the values have been determined, both samples have been reweighted to the p_T spectrum of the leading jet and the jet multiplicity of the data sample.

Model	Tune	$\epsilon_{4j}^{\mathrm{gen}}$	$\epsilon_{4j}^{\mathrm{det}}$
PYTHIA 8	`CUETP8M1`	0.4120	0.4039
HERWIG++	`CUETHS1`	0.4029	0.3923

For all 4-jet efficiencies, the statistical error is negligible. The residual difference between the 4-jet efficiencies obtained from the PYTHIA 8 and HERWIG++ samples, and the 4-jet efficiency determined from data is due to residual differences between the p_T spectra of the sub-leading and third leading jets of the inclusive single jet processes. However, the important result is the difference between the detector and generator level 4-jet efficiency. The deviation between the two levels is of sub-percent level and is an indication for an additional systematic uncertainty on the 4-jet efficiency. It is opted to use a more conservative error of 2% instead of an additional systematic uncertainty equal to 1%.

11.7 Systematic Uncertainty Summary

The total uncertainty can be calculated by taking the squared sum of all systematic uncertainties and the statistical uncertainty. The distinction between a lower and upper total uncertainty is made since the JES uncertainty is asymmetric. The systematic uncertainty is calculated for all observables in terms of their absolute cross section.

An overview of all systematic uncertainties is given in Tab. 11.2. Fig. 11.6, Fig. 11.7, Fig. 11.8 and Fig. 11.9 show the uncertainties visually for all observables in terms of the absolute cross section. The trigger, vertex selection and luminosity uncertainties are not shown in the plots. The latter is equal to 2.5% in all bins of all observables, while the two former uncertainties are of sub-percent level. The JES uncertainty is dominant for the observables in terms of the absolute cross section. Depending on the observable, significant contributions from the JER, model or statistical uncertainties can occur. The total upper and lower uncertainties never exceed 35%, except for the first bin of the p_T spectrum of the leading jet. Generally, the lower total uncertainty is smaller than the upper total uncertainty.

In order to perform a more qualitative study of the different models, the DPS-sensitive observables are normalized to one or the average of multiple bins in the region where the lowest DPS-sensitivity is expected. The different systematic uncertainties largely or completely cancel by performing the normalization, reducing the total uncertainty of the observables. Normalizing to the average of multiple bins aims to reduce the effect of statistical fluctuations, which is necessary for the $\Delta\phi_{\mathrm{soft}}$, $\Delta\phi_{\mathrm{3j}}^{\mathrm{min}}$ and $\Delta\mathrm{Y}$ observables. The $\Delta\phi_{\mathrm{soft}}$ observable has been normalized to its first five bins, covering the tail of the distribution, which is dominated by effects of the jet cone size. Jets that lie close to each other in the azimuthal plane of the detector can be reconstructed as one, effectively yielding the observed bump in the $\Delta\phi_{\mathrm{soft}}$ distribution. The $\Delta\phi_{\mathrm{3j}}^{\mathrm{min}}$ and $\Delta\mathrm{Y}$ observables are normalized to their first four bins. The calculation of the systematic uncertainties for the bin-normalized distributions is performed in the exact same manner as before. The resulting uncertainties are tabulated in Tab. 11.2 and are visually shown in Fig. 11.10. The normalization has to most profound effect on the JES uncertainty. The dominant uncertainty has become the statistical uncertainty for most of the observables.

Table 11.2: An overview of systematic uncertainties of all the observables, along with the statistical and the total uncertainties. All uncertainties are given in percentages. The JES uncertainty leads to asymmetric errors, therefore, an upper and lower uncertainty is tabulated. All other systematic uncertainties and the statistical uncertainty are symmetric. The systematic uncertainties were determined for all observables in terms of their absolute cross section and for the bin-normalized distributions. Due to the division in the bin-normalized cross sections, the systematic uncertainties partially or completely cancel, i.e. the ϕ_{ij}, $\Delta p_{\mathrm{T,soft}}$ and $\Delta\mathrm{S}$ distributions have systematic uncertainties equal to zero in the bin the histogram is normalized to. The systematic uncertainties for the inclusive single jet cross sections, the $\Delta\mathrm{S}_{\mathrm{DPS}}$ distribution and the 4-jet efficiency, which are needed in for the extraction of the σ_{eff}, are added as well. The 4-jet efficiency ϵ_{4j} has no JER or model uncertainty as it is defined as a ratio on detector level. The additional uncertainty on the ϵ_{4j} due to possible differences between generator and detector level is 2%, which is taken into account in the calculation of the total systematic uncertainties.

Observable	JES		JER	Model	Trigger	Vertex	Lum.	Stat.	Total	
	Upper	Lower							Upper	Lower
Absolute Cross Section										
$p_{\mathrm{T,1st}}$	11-39	9-30	2-26	0-16	< 1	< 1	2.5	1-10	11-51	10-44
$p_{\mathrm{T,2nd}}$	11-31	10-24	0-2	0-7	< 1	< 1	2.5	1-8	14-33	11-26
$p_{\mathrm{T,3rd}}$	1-31	7-24	1-3	2-7	< 1	< 1	2.5	2-15	13-33	13-25
$p_{\mathrm{T,4th}}$	10-25	0-21	1-8	2-7	< 1	< 1	2.5	4-31	14-34	13-32
η_1	22-33	18-28	< 1	1-9	< 1	< 1	2.5	3-5	22-34	19-29
η_2	22-30	18-26	< 1	0-6	< 1	< 1	2.5	3-6	23-31	18-26
η_3	21-29	18-24	< 1	0-7	< 1	< 1	2.5	3-5	22-30	19-25
η_4	19-29	16-24	< 1	1-8	< 1	< 1	2.5	3-4	19-30	17-25
$\Delta\phi_{\mathrm{soft}}$	21-24	19-20	< 1	1-7	< 1	< 1	2.5	3-4	22-25	20-22
$\Delta\phi_{\mathrm{3j}}^{\mathrm{min}}$	21-28	18-24	< 1	1-6	< 1	< 1	2.5	3-7	21-29	19-25
$\Delta\mathrm{Y}$	22-25	16-33	< 1	0-6	< 1	< 1	2.5	3-6	23-26	17-34
ϕ_{ij}	23-26	19-22	< 1	0-7	< 1	< 1	2.5	3-4	24-27	19-22
$\Delta p_{\mathrm{T,soft}}$	22-25	19-20	0-3	2-6	< 1	< 1	2.5	3-4	23-26	19-21
$\Delta\mathrm{S}$	4-34	13-20	< 1	0-5	< 1	< 1	2.5	3-13	12-37	15-22
Bin-normalized Cross Section										
$\Delta\phi_{\mathrm{soft}}$	0-1	0-1	< 1	0-4	< 1	< 1	-	3-4	3-6	3-6
$\Delta\phi_{\mathrm{3j}}^{\mathrm{min}}$	0-5	0-4	< 1	0-4	< 1	< 1	-	3-7	4-8	3-8
$\Delta\mathrm{Y}$	0-2	0-18	< 1	0-5	< 1	< 1	-	3-6	3-10	3-21
ϕ_{ij}	0-3	0-2	< 1	0-4	< 1	< 1	-	3-4	3-6	3-6
$\Delta p_{\mathrm{T,soft}}$	0-2	0-2	0-2	0-2	< 1	< 1	-	3-4	3-5	3-5
$\Delta\mathrm{S}$	0-16	0-7	< 1	0-7	< 1	< 1	-	3-13	3-22	3-15
σ_{eff} Extraction										
ϵ_{4j}	11	20	-	-	< 1	< 1	-	< 1	11	20
η_{A}	7-28	6-24	< 1	0-6	< 1	< 1	2.5	2-4	9-29	8-25
η_{B}	1-11	1-14	< 1	0-6	< 1	< 1	2.5	2-3	5-11	5-24
$\Delta\mathrm{S}_{\mathrm{DPS}}$	7-19	15-24	< 1	0-3	< 1	< 1	2.5	1-2	7-19	15-25

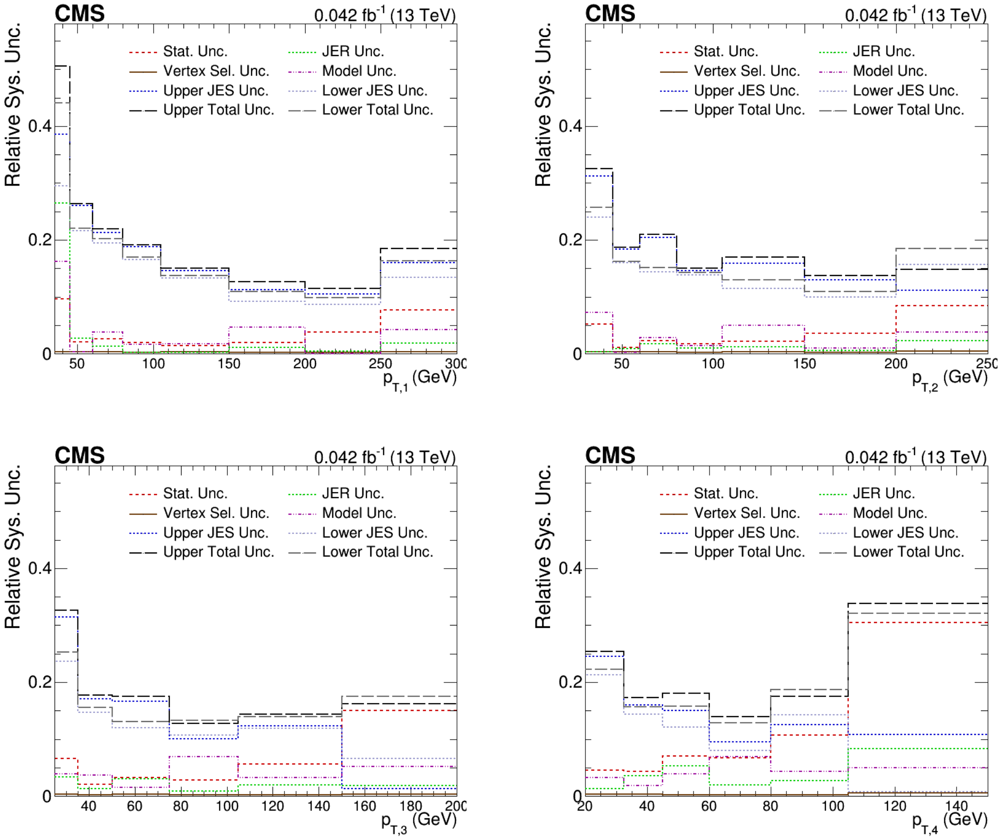


Figure 11.6: Visual representation of all systematic uncertainties of the p_T spectrum for the leading (top right), sub-leading (top left), third leading (bottom right) and last leading (bottom left) jets is shown. The trigger, vertex selection and luminosity uncertainties are left out for clarity's sake.

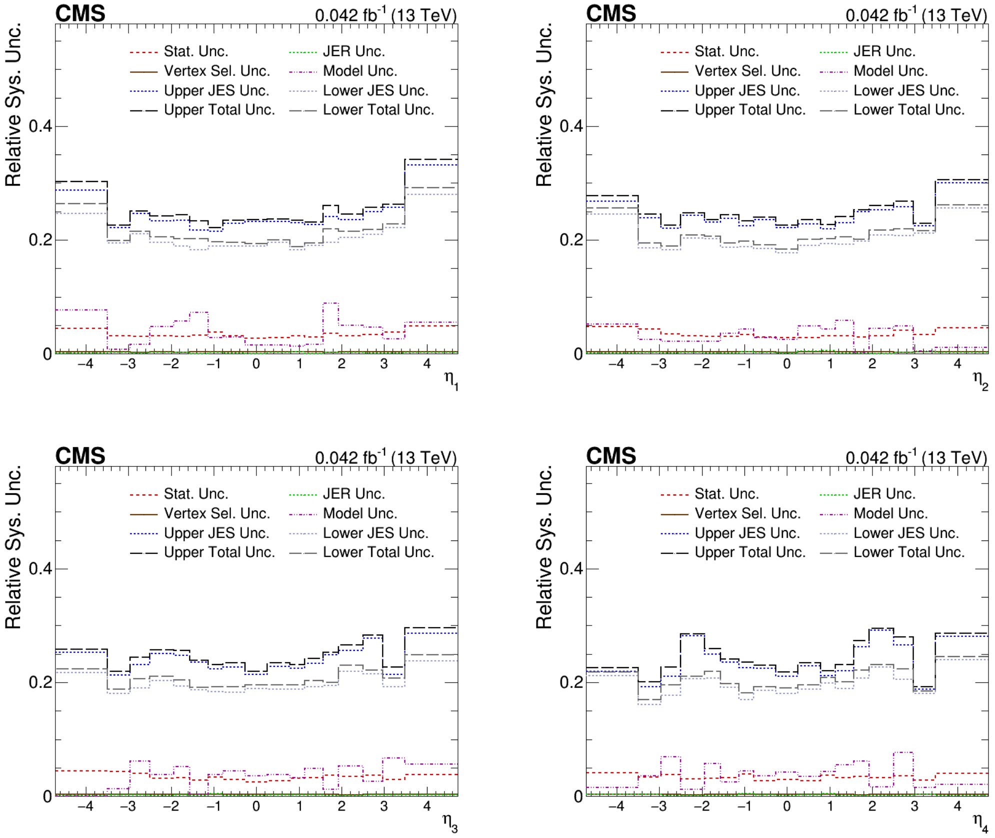


Figure 11.7: Visual representation of all systematic uncertainties of the η spectrum for the leading (top right), sub-leading (top left), third leading (bottom right) and last leading (bottom left) jets is shown. The trigger, vertex selection and luminosity uncertainties are left out for clarity's sake.

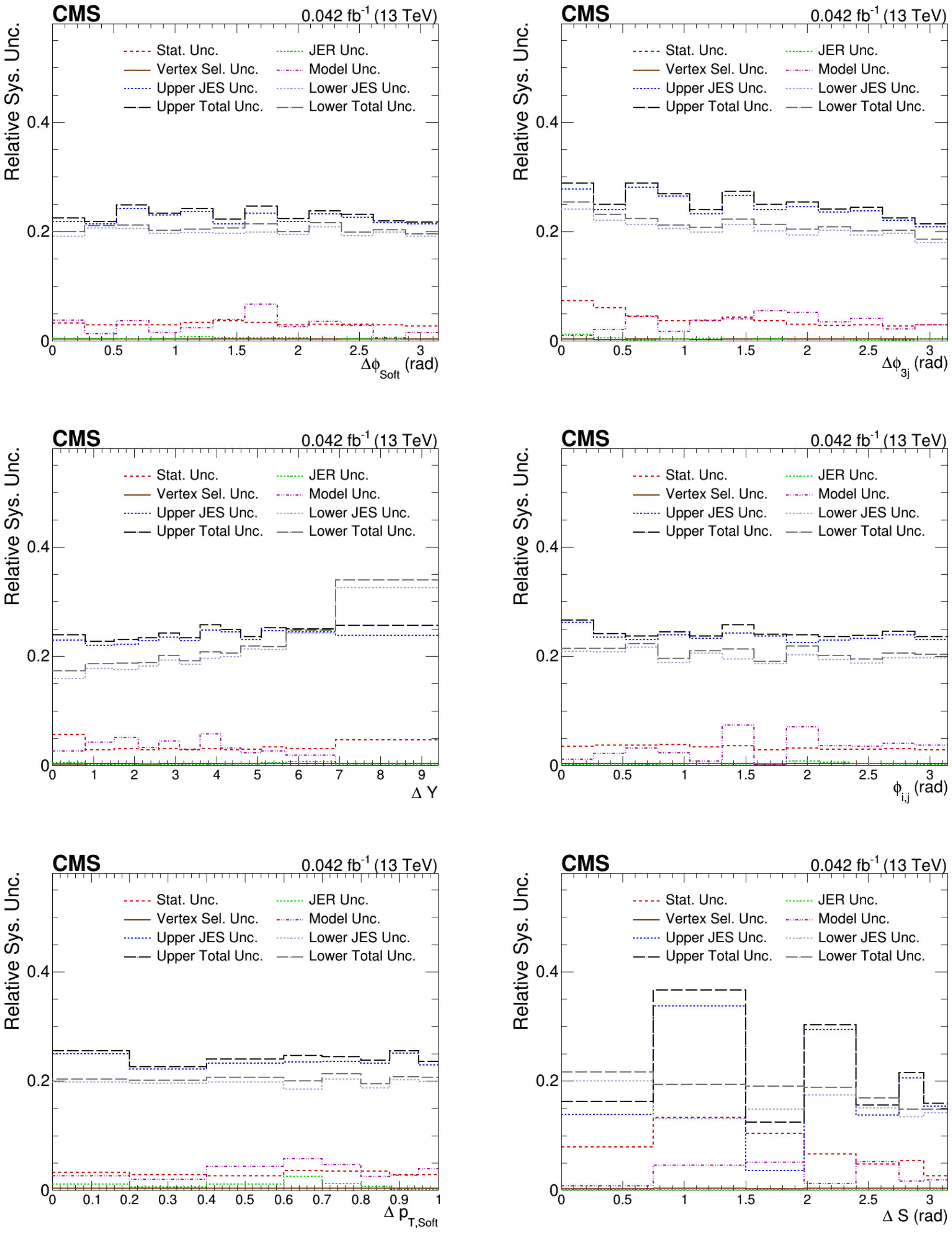


Figure 11.8: Visual representation of all systematic uncertainties of the DPS-sensitive observables is shown. The trigger, vertex selection and luminosity uncertainties are left out for clarity's sake.

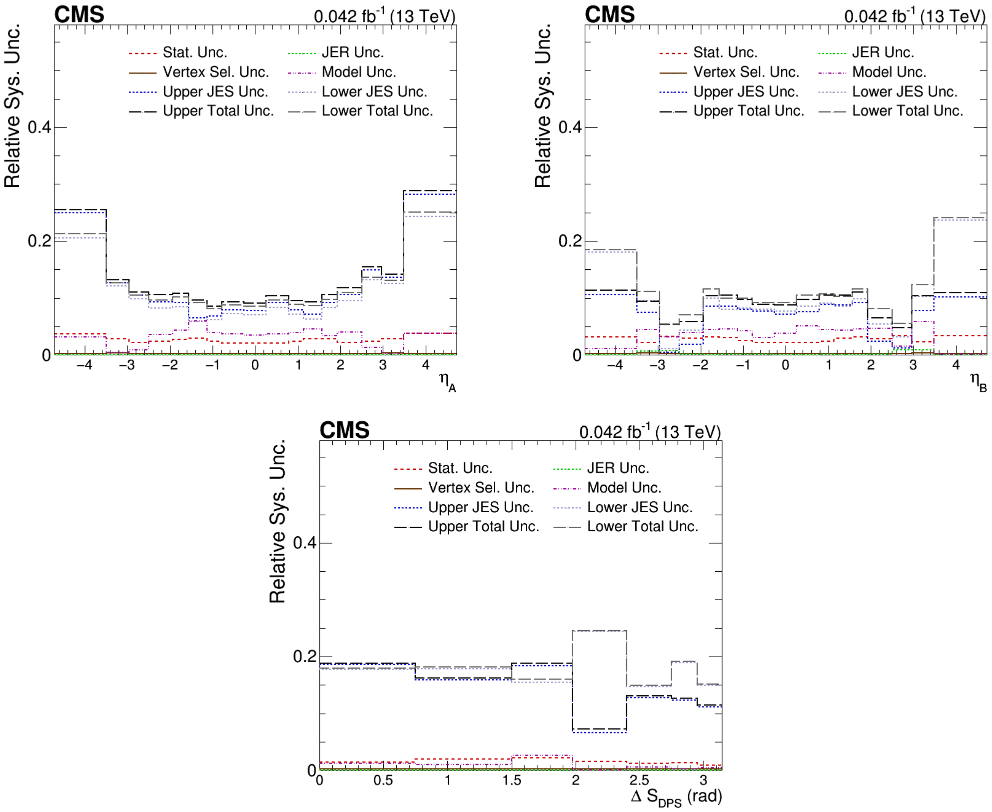


Figure 11.9: Visual representation of all systematic uncertainties of the η spectra of the inclusive single jet processes and the ΔS_{DPS} observable is shown. The trigger, vertex selection and luminosity uncertainties are left out for clarity's sake.

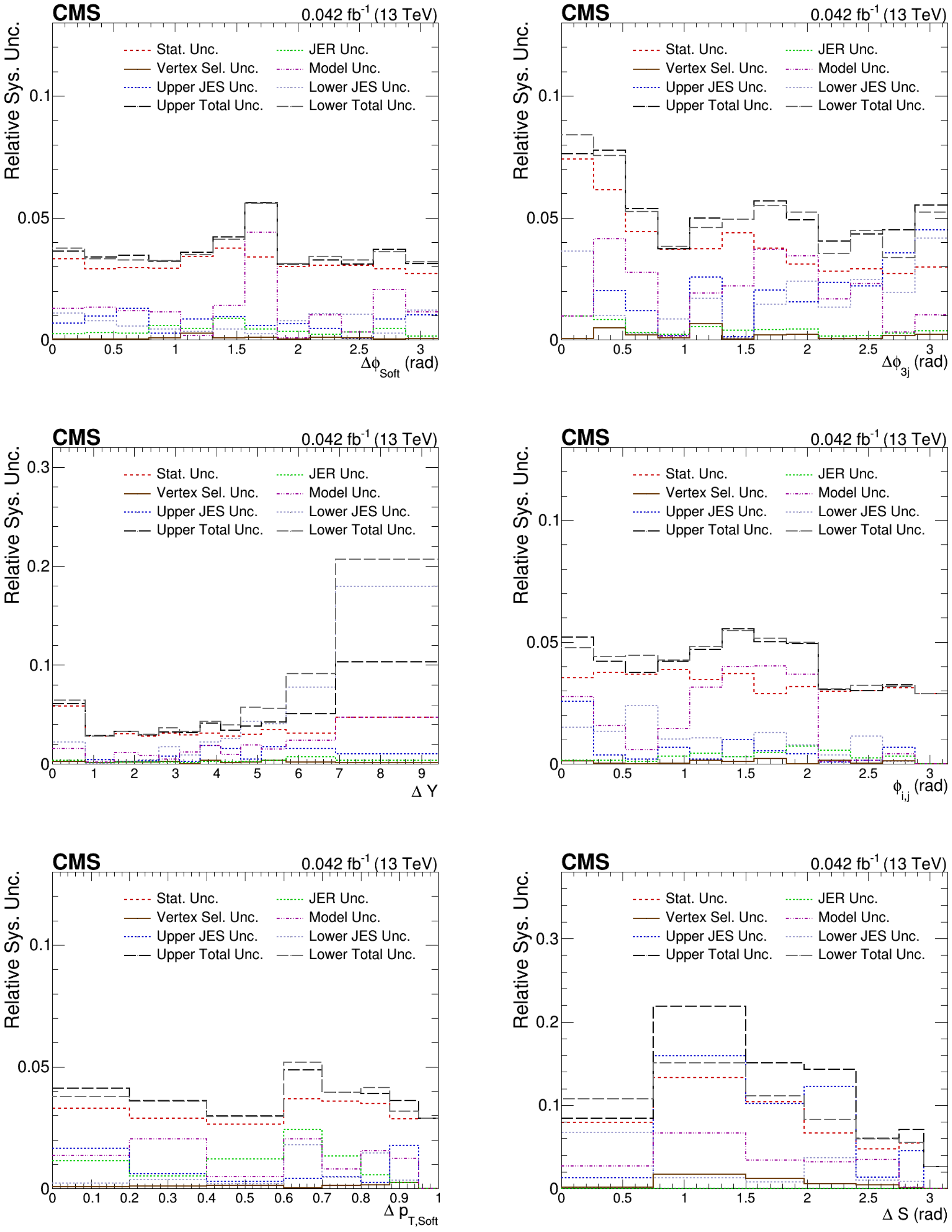


Figure 11.10: Visual representation of all systematic uncertainties of the bin-normalized DPS-sensitive observables is shown. The trigger, vertex selection and luminosity uncertainties are left out for clarity's sake.

Chapter 12

Results

The observables have been corrected to generator level, and their respective systematic uncertainties have been determined. Next, the cross section of the inclusive four jet production in both *region I* and *region II* will be determined, along with the cross sections of the inclusive single jet processes. Afterwards, the data distributions will be compared to the models introduced in Sec. 3.7. The models are divided in the three classes as before. The PYTHIA 8 and HERWIG models are all based on a LO 2 → 2 ME combined with different tunes, while the multijet models have implemented LO 2 → n with $n > 2$, NLO 2 → 2 or NLO 2 → 3 ME. The third and last class are the so-called SPS+DPS samples. They include the SPS models with an explicit DPS contribution, simulated by overlaying two 2 → 2 ME in one event. The final section of the chapter will use the different SPS models in order to extract the fraction of DPS events f_{DPS}, the DPS cross section $\sigma^{\mathrm{DPS}}_{\mathrm{A,B}}$ and the effective cross section σ_{eff} by applying the techniques laid out in Chap. 8.

12.1 Determination of the Cross Section

The different cross sections are extracted from the corresponding pseudorapidity spectra of the leading jet, except for the inclusive four jet cross section in the phase space domain *region II*, where the ΔS observable has been used. The systematical uncertainties are determined by varying the distributions accordingly and calculating the varied cross sections, while taking correlations into account. The squared sum of the discrepancies between the varied and nominal cross sections gives the total systematical uncertainty.

The cross sections for the inclusive four jet production in the phase space domains *region I* and *region II* are given below, respectively.

$$\sigma_{\mathrm{I}}\,(\mathrm{pp} \to 4\mathrm{j} + \mathrm{X}) = 2.77 \pm 0.02\ (\mathrm{stat.})\ ^{+0.68}_{-0.55}\ (\mathrm{syst.})\ \mu\mathrm{b}, \tag{12.1}$$

$$\sigma_{\mathrm{II}}\,(\mathrm{pp} \to 4\mathrm{j} + \mathrm{X}) = 0.61 \pm 0.01\ (\mathrm{stat.})\ ^{+0.12}_{-0.10}\ (\mathrm{syst.})\ \mu\mathrm{b}. \tag{12.2}$$

Tab. 12.1, Tab. 12.2 and Tab. 12.3 compare the values of the cross sections measured from data with the cross sections for all different models.

The cross sections for the inclusive single jet processes A and B are shown below.

$$\sigma_{\mathrm{A}} = 15.9 \pm 0.1 \text{ (stat.) } {}^{+1.8}_{-1.6} \text{ (syst.) } \mu\mathrm{b} \tag{12.3}$$

$$\sigma_{\mathrm{B}} = 106 \pm 1 \text{ (stat.) } {}^{+10}_{-11} \text{ (syst.) } \mu\mathrm{b} \tag{12.4}$$

The cross sections will be used as an input in the DPS pocket formula together with the 4-jet efficiency $\epsilon_{4\mathrm{j}} = 0.324\ {}^{+0.037}_{-0.065}$ (syst.) for the determination of f_{DPS}, $\sigma^{\mathrm{DPS}}_{\mathrm{A,B}}$ and σ_{eff}. The statistical uncertainty on the 4-jet efficiency is negligible compared to the systematical uncertainty.

12.2 PYTHIA 8 and HERWIG Models

PYTHIA 8, HERWIG++ and HERWIG 7 are the three main generators used for the models based on a LO 2 → 2 ME. The event generators will abbreviated as P8, H++ and H7 in the legends of the plots. Not all model predictions will be shown in terms of the absolute cross section for clarity's sake. However, in the ratio plots all models will be shown, and the distinction will be made between models with a p_{T}-ordered shower versus an angular-ordered/dipole-antenna shower. All details regarding the event generators and the used tunes can be found in Sec. 3.7.1. An overview of the models is shown in Tab. 12.1 along with their respective cross sections for both phase space domains *region I* and *region II*. All PYTHIA 8 and HERWIG models overestimate the cross sections compared to those measured from the data, especially PYTHIA 8 using the `CDPSTP8S1-4j` tune, which predicts cross sections roughly three times as large.

The p_{T} spectra in Fig. 12.1 show an abundance of low p_{T} jets, resulting in a much larger cross section. At p_{T} values larger as ~ 100 GeV, the models show agreement within 50% of the data and even within the total uncertainty for the HERWIG 7 models. From the pseudorapidity spectra, shown in Fig. 12.2, one may conclude that the excess in low p_{T} jets is unevenly distributed. The ratio of the cross sections rises in the forward and backward regions.

Table 12.1: Cross section measurements for the data and all the PYTHIA 8 and HERWIG models. The second cross section is for the variable ΔS specifically, as the variable is measured in *region II*, compared to *region I* for all other observables.

Sample	ME	Tune	PDF	σ_{I} (μb)	σ_{II} (μb)
Data	-	-	-	$2.77 \pm 0.02\ {}^{+0.68}_{-0.55}$	$0.61 \pm 0.01\ {}^{+0.12}_{-0.10}$
PYTHIA 8	LO 2 → 2	`CUETP8M1`	`NNPDF2.3 LO`	5.03	1.07
PYTHIA 8	LO 2 → 2	`CP5`	`NNPDF3.1 nnlo`	4.07	0.84
PYTHIA 8	LO 2 → 2	`CDPSTP8S1-4j`	`CTEQ6L1`	7.06	1.28
PYTHIA 8 +VINCIA	LO 2 → 2	Standard PYTHIA 8.3	`NNPDF2.3 LO`	4.66	0.97
HERWIG++	LO 2 → 2	`CUETHS1`	`CTEQ6L1`	4.35	0.83
HERWIG 7	LO 2 → 2	`CH3`	`NNPDF3.1 nnlo`	4.82	0.98
HERWIG 7	LO 2 → 2	`SoftTune`	`MMHT2014lo68cl`	5.34	1.07

The distributions of the DPS-sensitive observables are shown in Fig. 12.3, Fig. 12.4 and Fig. 12.5. The distributions on the left are in terms of the absolute cross section, while the distributions on the right have been normalized to one or multiple bins where a much reduced DPS-sensitivity is expected. The exact normalization is detailed in Sec. 11.7. The same large discrepancy between the cross section of the data and the PYTHIA 8 and HERWIG models becomes immediately apparent just as in Tab. 12.1.

The shape of the $\Delta\phi_{\mathrm{soft}}$ and $\Delta p_{\mathrm{T,soft}}$ observables in Fig. 12.3 and 12.5, respectively, is relatively well described by all PYTHIA 8 and HERWIG models compared to other the other observables. No deviations larger than 20% are found, albeit being larger than the total uncertainty on the data points.

When examining the $\Delta\phi_{\mathrm{3j}}^{\mathrm{min}}$ and ϕ_{ij} observables, a clear distinction can be made between the models interfaced with a p_{T}-ordered shower compared to the models interfaced with an angular-ordered/dipole-antenna shower. The former gives an accurate description of the $\Delta\phi_{\mathrm{3j}}^{\mathrm{min}}$ observable, but overshoot the DPS-sensitive tail of the ϕ_{ij} distribution. The latter models are too correlated in the DPS-sensitive peak of the $\Delta\phi_{\mathrm{3j}}^{\mathrm{min}}$ observable, while a good agreement is found with the shape of the ϕ_{ij} observable. The predictions made by PYTHIA 8 interfaced with VINCIA confirm that the parton shower lies at the base of the distinct behavior between the two groups of models. The sensitivity to the type of parton shower of both observables makes them less suitable to study DPS and to perform a possible extraction of the DPS cross section.

The PYTHIA 8 and HERWIG models all fail to describe the shape of the $\Delta\mathrm{Y}$ observable. The predictions are far too decorrelated in the DPS-sensitive tail since the ratio rises when going to higher values of $\Delta\mathrm{Y}$. The overshoot is due to the abundance of low p_{T} jets in the forward and backward pseudorapidity regions of the models compared to the data.

The predictions of the $\Delta\mathrm{S}$ observable exhibit the most consistent behavior for all models. The observable shows no large dependence on the type of parton shower. Most models undershoot the slope going to zero, leaving room for a DPS contribution to improve the description of the shape. Only the PYTHIA 8 models using the `CUETP8M1` and the DPS tune `CDPSTP8S1-4j` overshoot the slope, indicating a possible overestimation of the DPS contribution. The former is an older tune, while the latter is expected to perform well in the DPS-sensitive regions.

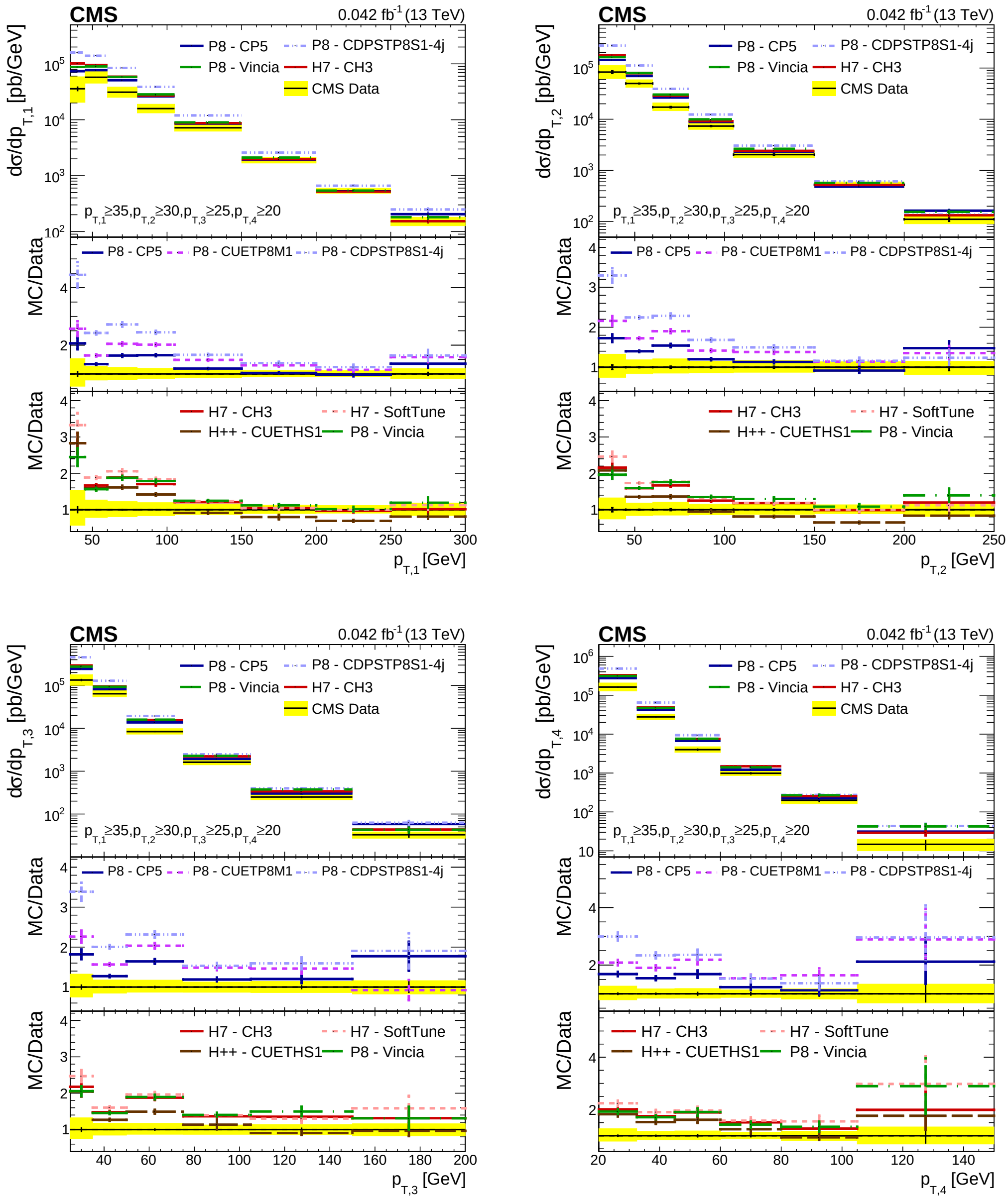


Figure 12.1: Comparison of the unfolded p_{T} spectra of the data with the PYTHIA 8 (P8), HERWIG++ (H++) and HERWIG 7 (H7) models for the leading (top left), sub-leading (top right), third leading (bottom left), and fourth leading (bottom right) jet. The error bars represent the statistical uncertainties, and the yellow band indicates the total uncertainty on the measurement.

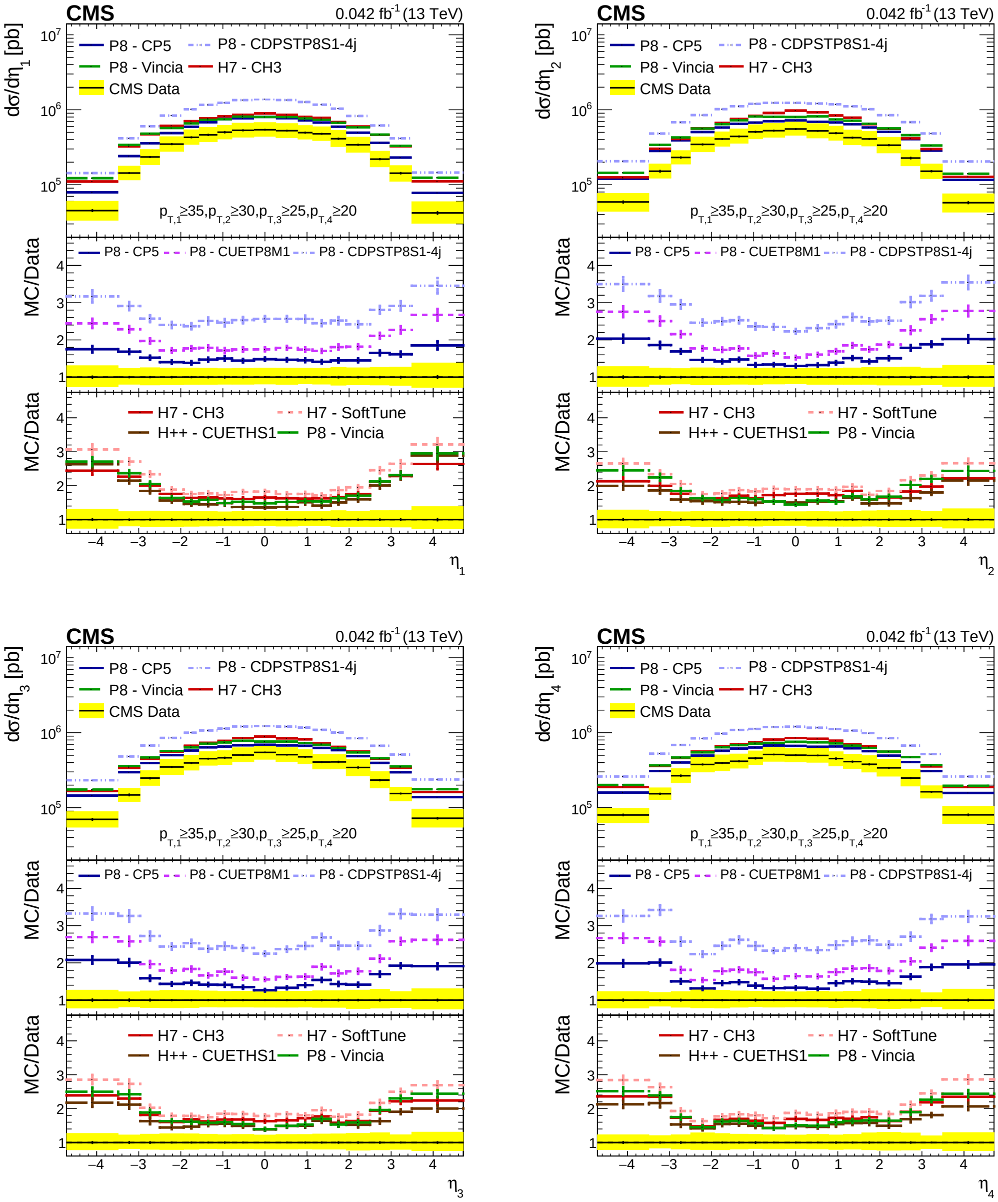


Figure 12.2: Comparison of the unfolded η spectra of the data with the PYTHIA 8 (P8), HERWIG++ (H++) and HERWIG 7 (H7) models for the leading (top left), sub-leading (top right), third leading (bottom left), and fourth leading (bottom right) jet. The error bars and bands are shown similarly to Fig. 12.1.

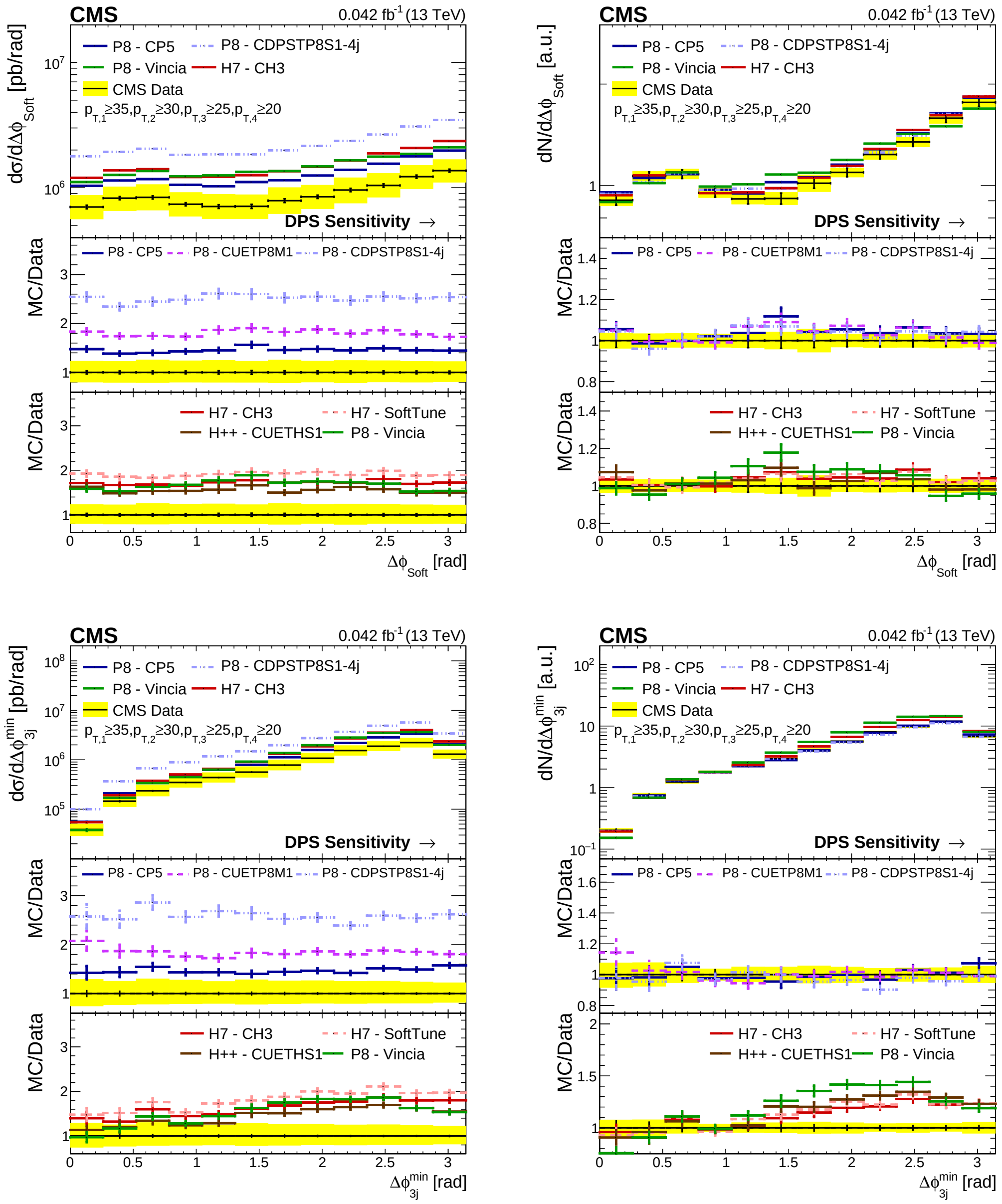


Figure 12.3: Comparison of the unfolded $\Delta\phi_{\text{soft}}$ and $\Delta\phi^{\text{min}}_{3\text{j}}$ observables of the data with the PYTHIA 8 (P8), HERWIG++ (H++) and HERWIG 7 (H7) models. The left-hand plots are in terms of the absolute cross section, while the right-hand plots are normalized to the region where reduced DPS-sensitivity is expected. The error bars and bands are shown similarly to Fig. 12.1.

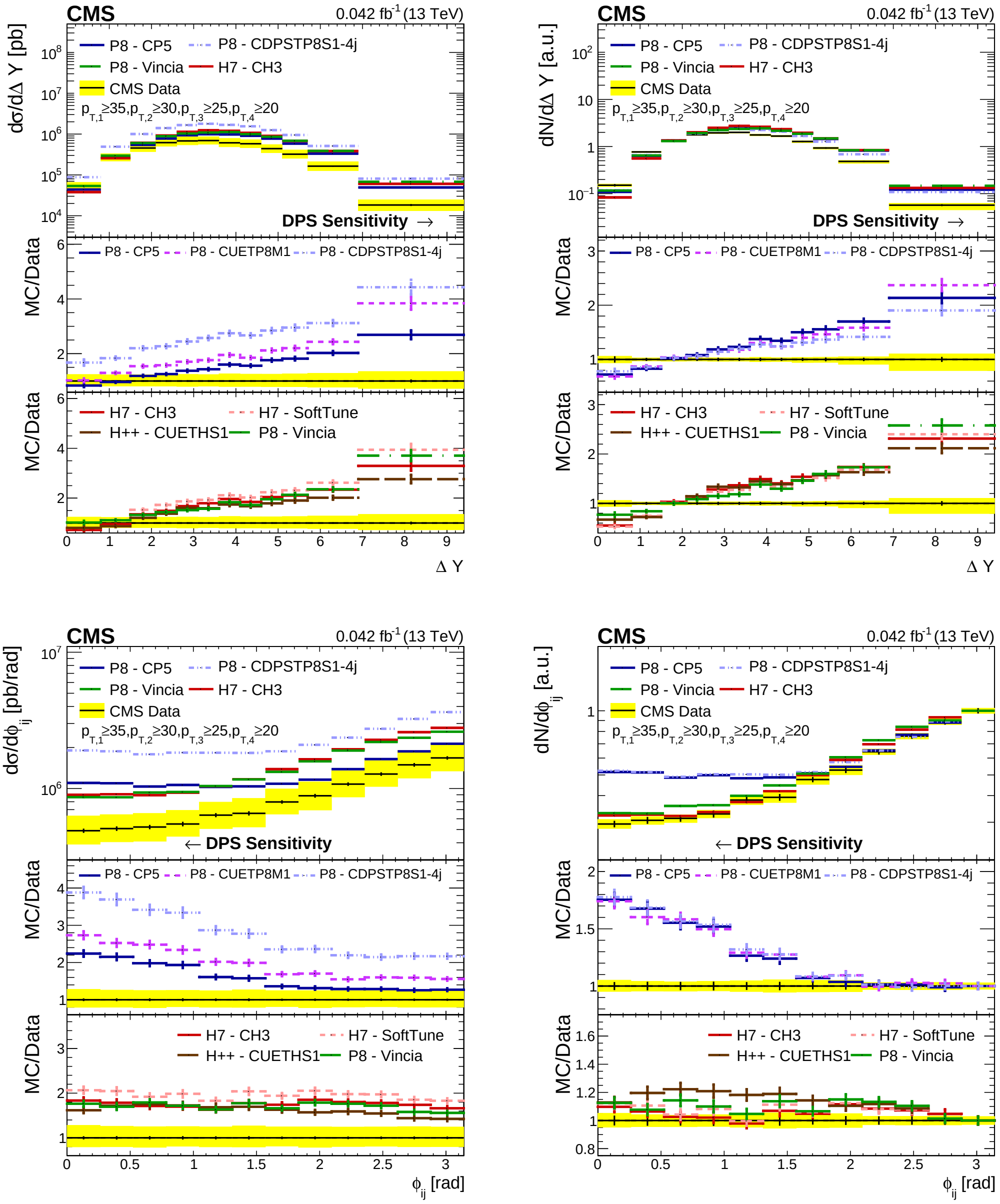


Figure 12.4: Comparison of the unfolded ΔY and ϕ_{ij} observables of the data with the PYTHIA 8 (P8), HERWIG++ (H++) and HERWIG 7 (H7) models. The left-hand plots are in terms of the absolute cross section, while the right-hand plots are normalized to the region where reduced DPS-sensitivity is expected. The error bars and bands are shown similarly to Fig. 12.1.

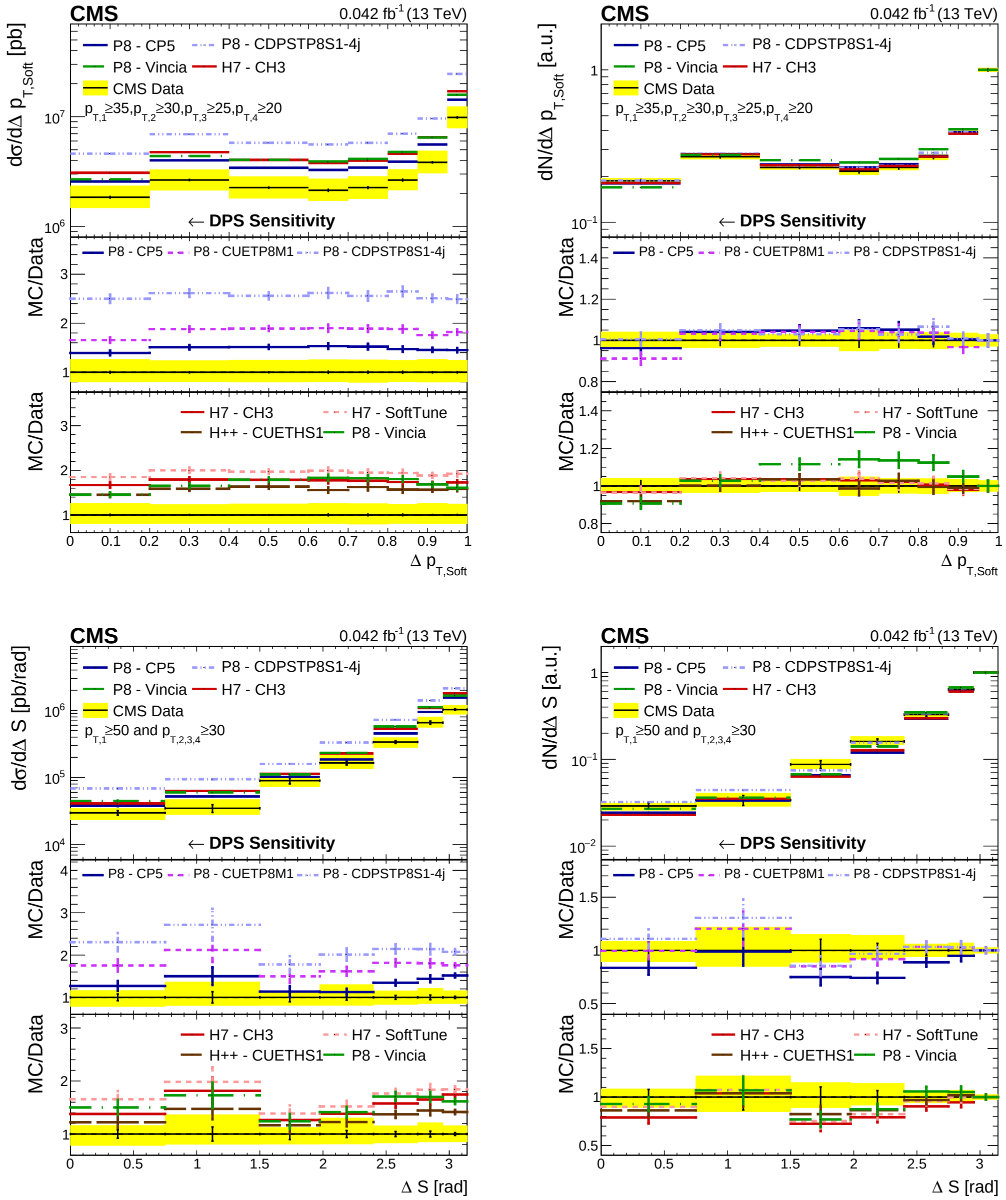


Figure 12.5: Comparison of the unfolded $\Delta p_{\mathrm{T,soft}}$ and ΔS observables of the data with the PYTHIA 8 (P8), HERWIG++ (H++) and HERWIG 7 (H7) models. The left-hand plots are in terms of the absolute cross section, while the right-hand plots are normalized to the region where reduced DPS-sensitivity is expected. The observable ΔS has been determined in *region II*, compared to *region I* for all other observables. The error bars and bands are shown similarly to Fig. 12.1.

12.3 Multijet Models

The second group of models are the so-called multijet models. They all have higher order ME at LO and/or NLO corrections implemented. Three different event generators have been used to produce predictions. Four samples have been generated by the KaTie event generator at LO with a $2 \to 4$ ME, two on-shell and two off-shell samples. Three sets of predictions were produced with the MadGraph 5 event generator, two LO predictions effectively combining a $2 \to 2$, a $2 \to 3$ and a $2 \to 4$, and one sample with NLO corrections using a $2 \to 2$ ME. The combined matrix element will be denoted as $2 \to 2, 3, 4$ from hence on forth. The powheg event generator is used for the last two sets of NLO predictions with a $2 \to 2$ and a $2 \to 3$ ME, respectively. The KaTie, MadGraph 5 and powheg event generators will be abbreviated in the legends of the distributions as KT, MG5 and PW, respectively, for clarity's sake. The type of tune or TMD will be added, allowing for a clear indication of all the models. Not all distributions will be shown in terms of the absolute cross section as there are simply too many. However, three different ratio plots for the KaTie, the LO $2 \to 2, 3, 4$ and NLO models will hold all the ratios. Further details and specifications of the models can be found in Sec. 3.7.2. In Tab. 12.2 an overview of all the MC samples with their respective tune or TMD along with their cross sections for both phase space domains *region I* and *region II* is given.

The cross sections of both the on-shell KaTie models interfaced with pythia 8 using the `CP5` and `CH3` overshoot the data and are up to three times as large as the cross sections obtained from data. The cross sections predicted by the off-shell KaTie models are significantly lower, where the cross section in *region I* shows agreement within the data uncertainty. Apart from the off-shell KaTie models, the MadGraph 5 LO sample interfaced with pythia 8 using the `CP5` tune and the powheg NLO $2 \to 3$ sample agree with the data within uncertainty. All other MadGraph 5 and powheg models predict cross sections much lower compared to the data.

The p_T spectra of the different models in comparison to the data are shown in Fig. 12.6. The on-shell KaTie models agree with the data in the first bin of each of the p_T spectra, however, they are all to hard when going to higher values of p_T, which is caused by the fact that all jets are most likely to originate from the $2 \to 4$ ME and not from the parton shower. The same effect, but less pronounced, can be seen for the off-shell KaTie curves. The different TMD sets for the off-shell KaTie models only result in small variations. A better description of the p_T spectra is given by the MadGraph 5 LO sample, with a p_T-ordered shower. Jets can originate from the parton shower, yielding a softer p_T spectrum. The combination of the MadGraph 5 LO $2 \to 2, 3, 4$ sample with the dipole-antenna showering from pythia 8+Vincia, results in a lowering of the total cross section. All NLO models give an analogue description compared to the MadGraph 5 sample, where the virtual corrections included in the higher-order ME can effectively lower the cross section. Comparing the multijet samples to the standard pythia 8 and herwig curves from the previous section, one can conclude that NLO corrections and the inclusion of multi-leg ME improve the description of the p_T spectra.

The pseudorapidity spectra are shown in Fig. 12.7. The central region is well described by all models, even the on-shell KaTie models. The overall cross section is to large but the ratio remains more or less constant for $|\eta| \leq 3.0$. An excess of jets in the forward and backward regions is be observed as the ratio rises, however, it is less pronounced than in the case of the pythia 8 and herwig models. The excess is also strongest for the leading jet and diminishes when going to the second, third and fourth leading jet, yielding a good description of the shape of the last leading jet.

Table 12.2: Cross section measurements for the data and all the multijet models. The second cross section is for the variable ΔS specifically, as the variable is measured in *region II*, compared to *region I* for all other observables.

Sample	ME	Tune	PDF/TMD	$\sigma_{\rm I}$ (μb)	$\sigma_{\rm II}$ (μb)
Data	-	-	-	$2.77 \pm 0.02\ ^{+0.68}_{-0.55}$	$0.61 \pm 0.01\ ^{+0.12}_{-0.10}$
KaTie on-shell + pythia 8	LO $2 \to 4$	`CP5`	`NNPDF3.1 nnlo`	4.23	2.87
KaTie on-shell + herwig 7	LO $2 \to 4$	`CH3`	`NNPDF3.1 nnlo`	3.56	2.25
KaTie off-shell + Cascade	LO $2 \to 4$	-	`MRW`	2.40	1.46
KaTie off-shell + Cascade	LO $2 \to 4$	-	`PBTMD`	2.57	1.56
MadGraph 5 + pythia 8	LO $2 \to 2, 3, 4$	`CP5`	`NNPDF3.1 nnlo`	2.69	1.26
MadGraph 5 + pythia 8+Vincia	LO $2 \to 2, 3, 4$	Standard pythia 8.3	`NNPDF2.1 LO`	1.93	0.90
MadGraph 5 + pythia 8	NLO $2 \to 2$	`CP5`	`NNPDF3.1 nnlo`	2.12	1.03
powheg + pythia 8	NLO $2 \to 2$	`CP5`	`NNPDF3.1 nnlo`	3.50	1.62
powheg + pythia 8	NLO $2 \to 3$	`CP5`	`NNPDF3.1 nnlo`	2.55	1.22

All differential cross sections of the DPS-sensitive observables are shown in Fig. 12.8, Fig. 12.9 and Fig. 12.10. The plots on the left show the distributions in terms of the absolute cross section, while the distributions on the right display the bin-normalized observables.

Let us first focus on the observables in terms of the absolute cross section. The observables $\Delta\phi_{\rm soft}$, $\Delta\phi_{\rm 3j}^{\rm min}$, $\Delta p_{\rm T,soft}$ and $\phi_{\rm ij}$ are described well by the off-shell KaTie models, the MadGraph 5 LO $2 \to 2, 3, 4$ with a $p_{\rm T}$-ordered shower and all powheg NLO models. The on-shell KaTie models, the MadGraph 5 LO $2 \to 2, 3, 4$ model with the dipole-antenna showering from Vincia and the MadGraph 5 NLO model predictions do not match the data. Their total cross sections are either too high or too low with respect to the total uncertainty of the measurement. The best description is given by the off-shell KaTie and powheg NLO $2 \to 3$ models. They agree within uncertainty with the data for these variables, except for the bins where a DPS contribution is expected. The MadGraph 5 LO $2 \to 2, 3, 4$ prediction with a $p_{\rm T}$-ordered shower, displays agreement with the data within the uncertainty even in the DPS-sensitive areas.

The off-shell KaTie predictions for the ΔY variable show good agreement with the data, except for the last bin. All other models fail to describe the data by either underestimating the cross section at low values of ΔY or overestimating it for large values of ΔY.

The ΔS observable is well described by the off-shell KaTie models and the MadGraph 5 LO $2 \to 2, 3, 4$ with a $p_{\rm T}$-ordered shower. The latter performs well throughout the whole domain, while the former two models perform well in the peak but less so in the slope of the distribution.

Switching to the bin-normalized distributions, the shape of the observables $\Delta\phi_{\mathrm{soft}}$ and $\Delta p_{\mathrm{T,soft}}$ deviates strong between the different models in the DPS-sensitive areas. The MadGraph 5 LO $2 \rightarrow 2, 3, 4$ samples tend to overestimate the data, while all other models underestimate the data, indicating a need for DPS.

The $\Delta\mathrm{Y}$ observable is only properly described by the off-shell KaTie models, except for the last bin, confirming the observation made earlier.

In contrast to the predictions of the $\Delta\phi_{3\mathrm{j}}^{\mathrm{min}}$ observable made by the pythia 8 and herwig models on the bottom left plot in Fig. 12.3, no clear distinction between the different types of parton shower can be made. The different models overestimate and underestimate the data indicating a more complex interplay between the order of the ME and the type of parton shower model. All models, except the MadGraph 5 predictions, show a sudden drop at higher values of $\Delta\phi_{3\mathrm{j}}^{\mathrm{min}}$, indicating the need for a DPS contribution.

When examining the bin-normalized distribution of the ϕ_{ij} observable, one does observe a correlation between how well the models describe the shape of the data and the used type of parton shower model. The description of the shape of the data for the on-shell KaTie models drastically improves when the angular-ordered shower from herwig 7 is used. Agreement between the herwig 7 model and the data is found within the uncertainty of the data. The off-shell KaTie samples with parton showers modeled by Cascade are the only models that undershoot the data and leave room for a DPS contribution in the tail of the variables ϕ_{ij}.

The observable $\Delta\mathrm{S}$ shows again a consistency as it did for the pythia 8 and herwig models, except for the KaTie models. They are the only predictions that overestimate the DPS-sensitive tail of the $\Delta\mathrm{S}$ distribution. The reason for the overestimation of the pedestal is linked to the p_{T} spectra. The four leading jets generally originate from the $2 \rightarrow 4$ ME, resulting in p_{T} spectra that are too hard. All other predictions undershoot the pedestal and leave room for a DPS contribution.

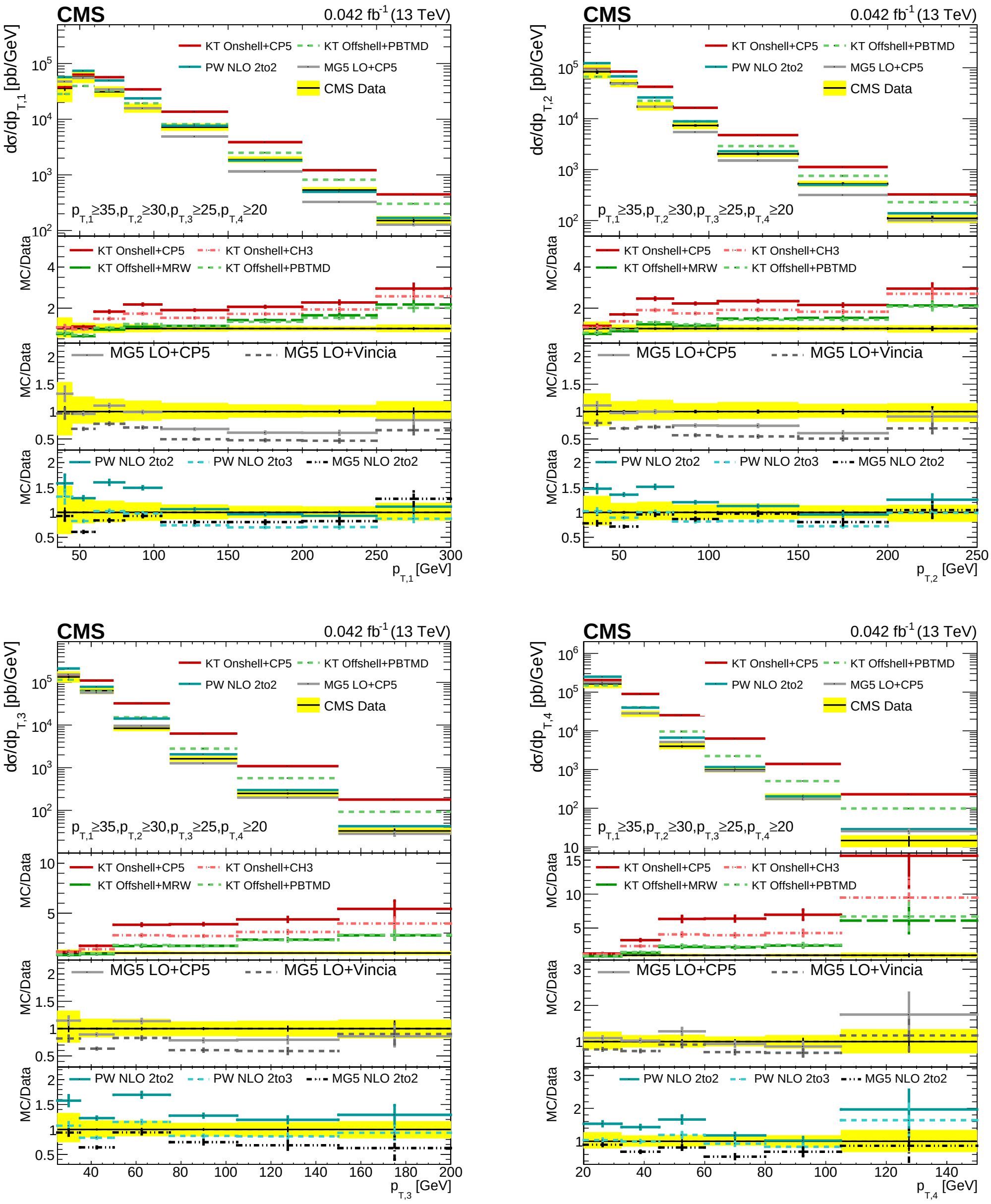


Figure 12.6: Comparison of the unfolded p_T spectra of the data with the KaTie (KT), MadGraph 5 (MG5) and powheg (PW) models for the leading (top left), sub-leading (top right), third leading (bottom left), and fourth leading (bottom right) jet. The error bars represent the statistical uncertainties, and the yellow band indicates the total uncertainty on the measurement.

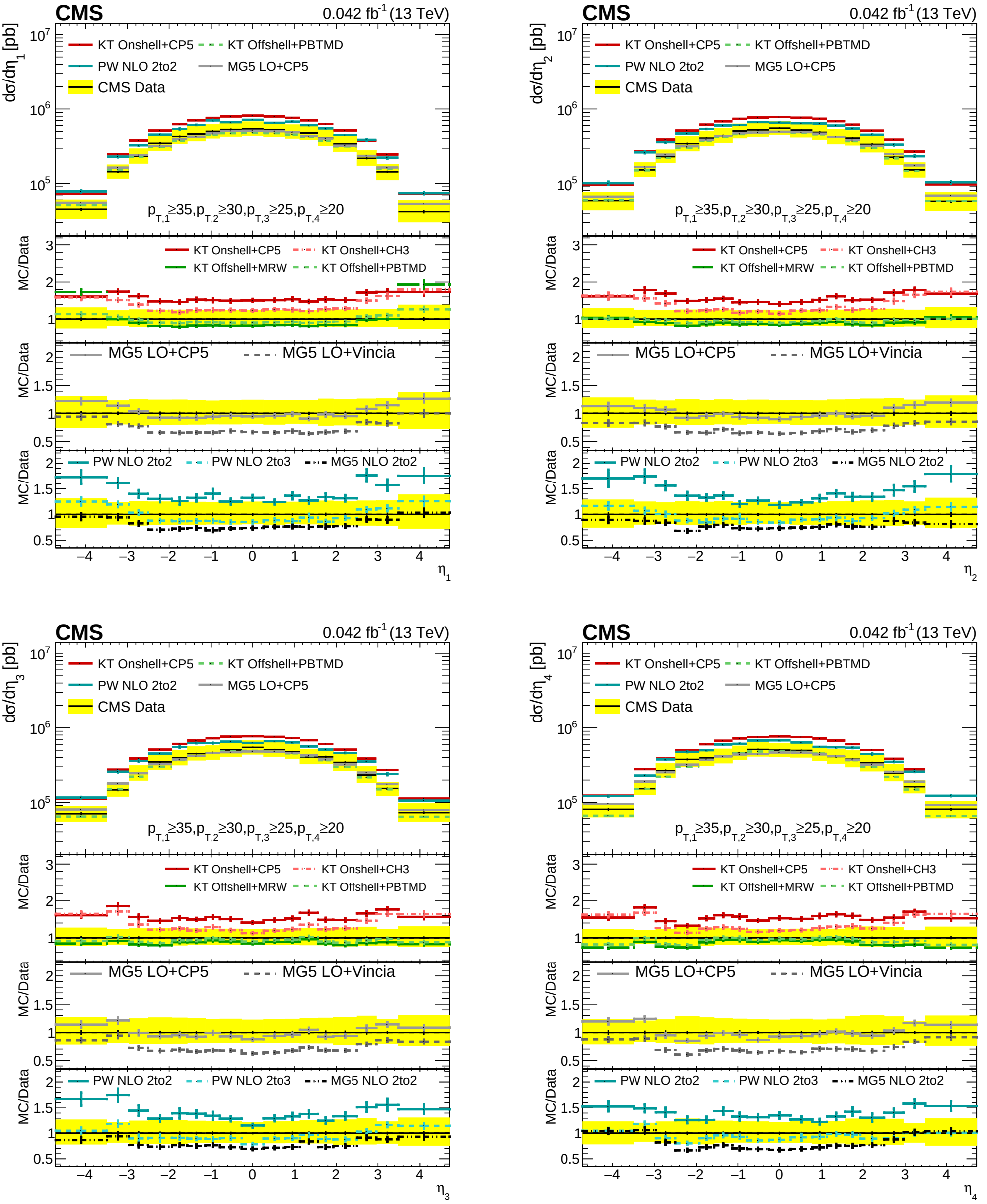


Figure 12.7: Comparison of the unfolded η spectra of the data with the KaTie (KT), MadGraph 5 (MG5) and powheg (PW) models for the leading (top left), sub-leading (top right), third leading (bottom left), and fourth leading (bottom right) jet. The error bars and bands are shown similarly to Fig. 12.6.

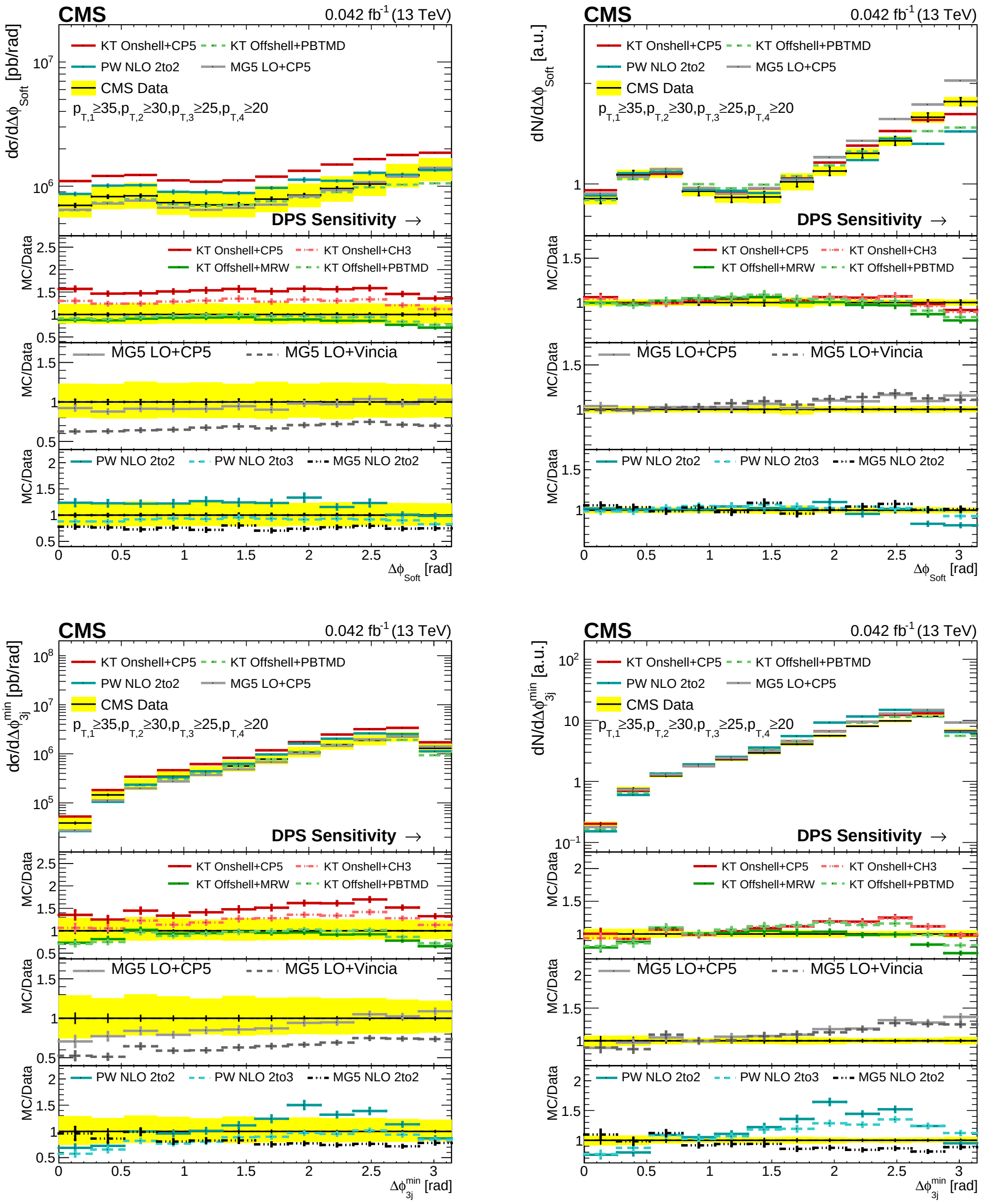


Figure 12.8: Comparison of the unfolded $\Delta\phi_{\text{soft}}$ and $\Delta\phi_{3\text{j}}^{\text{min}}$ observables of the data with the KATIE (KT), MADGRAPH 5 (MG5) and POWHEG (PW) models. The left-hand plots are in terms of the absolute cross section, while the right-hand plots are normalized to the region where reduced DPS-sensitivity is expected. The error bars and bands are shown similarly to Fig. 12.6.

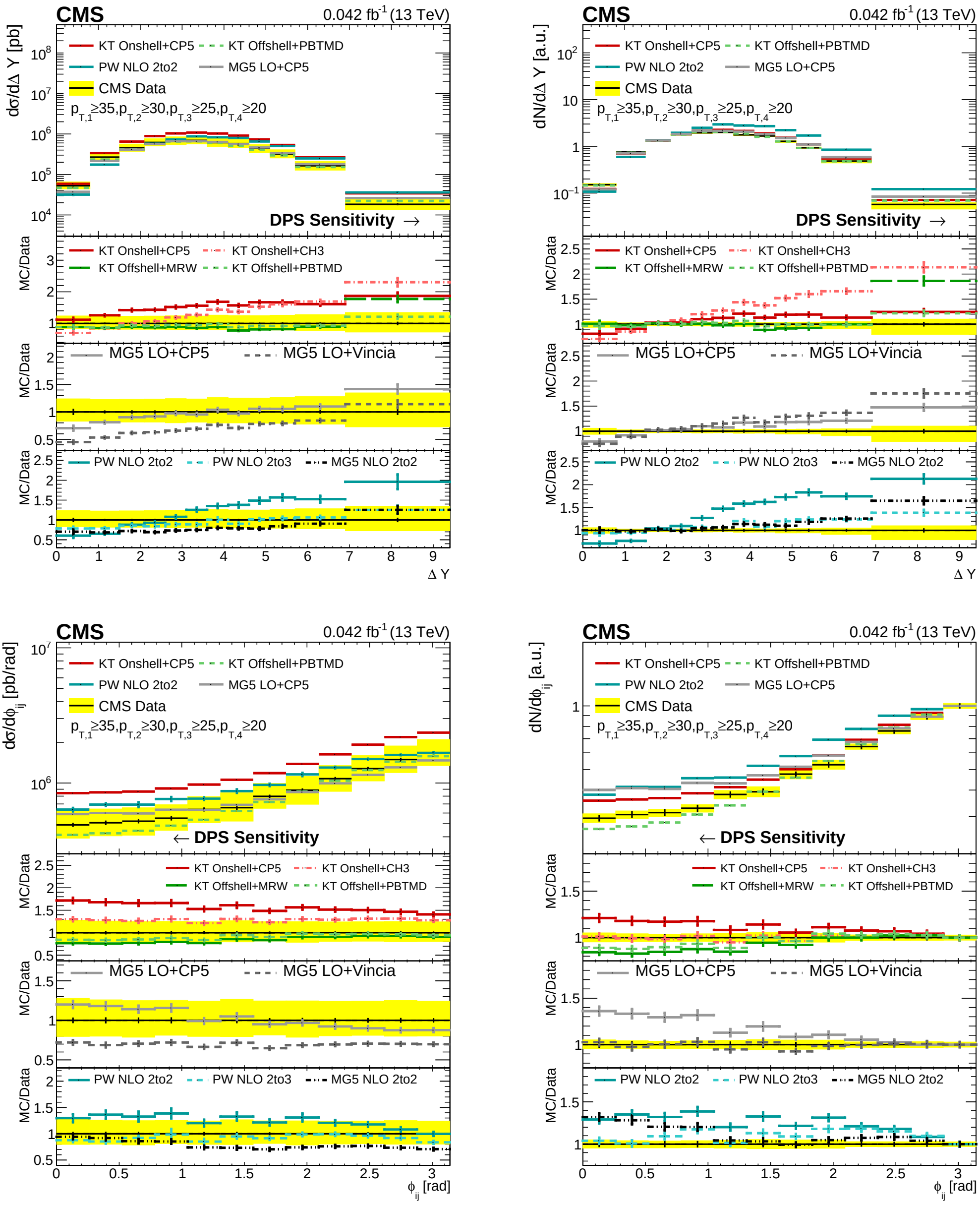


Figure 12.9: Comparison of the unfolded ΔY and ϕ_{ij} observables of the data with the KaTie (KT), MadGraph 5 (MG5) and powheg (PW) models. The left-hand plots are in terms of the absolute cross section, while the right-hand plots are normalized to the region where reduced DPS-sensitivity is expected. The error bars and bands are shown similarly to Fig. 12.6.

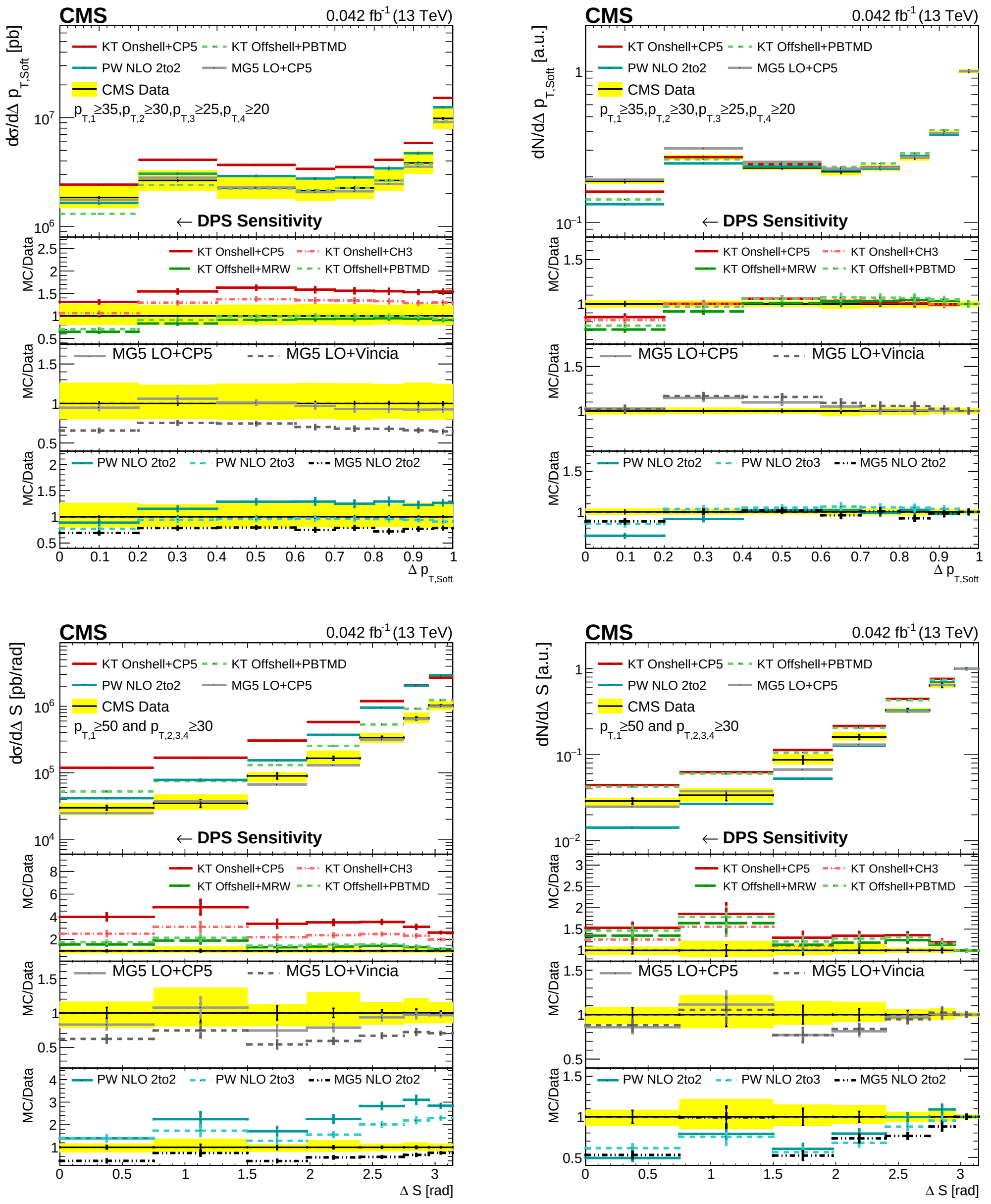


Figure 12.10: Comparison of the unfolded $\Delta p_{\mathrm{T,soft}}$ and ΔS observables of the data with the KaTie (KT), MadGraph 5 (MG5) and powheg (PW) models. The left-hand plots are in terms of the absolute cross section, while the right-hand plots are normalized to the region where reduced DPS-sensitivity is expected. The observable ΔS has been determined in *region II*, compared to *region I* for all other observables. The error bars and bands are shown similarly to Fig. 12.6.

12.4 SPS+DPS Models

The third and last group of models are the so-called SPS+DPS samples. All details can be found in Sec. 3.7.3. These predictions have been generated with the pythia 8 and KaTie MC event generators since both generators allow for the production of a pure DPS sample which is added to its SPS counterpart. Note that the predictions from the pythia 8 model using the `CDPSTP8S1-4j` tune are also included in this group of models and are the same compared to the predictions in Sec. 12.2. No explicit DPS contribution is needed in the model since the UE parameters have been specifically tuned to account for DPS contributions. In the legends of the figures, all models will be abbreviated in the exact same manner as before. An overview of all the SPS+DPS samples with their respective integrated cross sections is given in Tab. 12.3. Only the values of the off-shell KaTie models agree within the uncertainty with the measured cross section of the phase space domain *region I*.

Fig. 12.11 and Fig. 12.12 show the p_{T} and pseudorapidity spectra of the four leading jets, respectively. In comparison to the spectra of the pure SPS samples from the previous sections, one can conclude that the DPS samples mainly contribute in the low p_{T} region of the p_{T} spectra and mostly in the forward/backward regions of the pseudorapidity spectra.

Table 12.3: Cross section measurements for the data and all SPS+DPS models. The second cross section is for the ΔS observable specifically since the variable is measured in *region II*, compared to *region I* for all other observables.

Sample	ME	Tune	PDF/TMD	σ_{I} (μb)	σ_{II} (μb)
Data	-	-	-	$2.77 \pm 0.02\ ^{+0.68}_{-0.55}$	$0.61 \pm 0.01\ ^{+0.12}_{-0.10}$
SPS+DPS KaTie on-shell + pythia 8	LO 2 → 4	`CP5`	`NNPDF3.1 nnlo`	5.04	2.14
SPS+DPS KaTie off-shell + Cascade	LO 2 → 4	-	`MRW`	3.11	0.95
SPS+DPS KaTie off-shell + Cascade	LO 2 → 4	-	`PBTMD`	3.12	0.99
SPS+DPS pythia 8	LO 2 → 2	`CP5`	`NNPDF3.1 nnlo`	4.76	0.94
pythia 8	LO 2 → 2	`CDPSTP8S1-4j`	`CTEQ6L1`	7.06	1.28

All other variables are shown in Fig. 12.13 and Fig. 12.14. The observables on the left show the distributions in terms of the absolute cross sections, while the bin-normalized observables are displayed on the right. The predictions made by the off-shell KaTie samples mostly agree with the data within the uncertainty for all the observables, except for ΔS. However, excesses are observed compared to the data in the DPS-sensitive regions of the observables, while the pure SPS predictions underestimate the same regions in Fig. 12.8 and Fig. 12.9. The results indicate the need for a DPS contribution for a proper description of the data. However, a value for σ_{eff} equal to 21.3 mb, taken from [121], is too small for these models.

The bin-normalized distributions are shown on the right in Fig. 12.13 and Fig. 12.14. Only small increases of at most $\sim 5\%$ are observed in the DPS-sensitive regions when laying the pure SPS pythia 8 model using the `CP5` tune from Fig. 12.3, Fig. 12.4 and Fig. 12.5 next to its SPS+DPS counterpart. The increase in the on-shell and off-shell KaTie curves is more noticeable and is the largest for the off-shell samples.

Both SPS+DPS PYTHIA 8 samples give a good shape description of the observables $\Delta\phi_{\mathrm{soft}}$, $\Delta\phi_{3\mathrm{j}}^{\min}$, $\Delta p_{\mathrm{T,soft}}$ and $\Delta\mathrm{S}$. The `CDPSTP8S1-4j` tune performs slightly better compared to the `CP5` tune. These results might have been expected as the `CDPSTP8S1-4j` tune was obtained by fitting a PYTHIA 8 sample to the $\Delta\phi_{\mathrm{soft}}$, $\Delta p_{\mathrm{T,soft}}$ and $\Delta\mathrm{S}$ observables measured from data taken at a centre-of-mass energy of 7 TeV, as detailed in [121].

The pedestal of the $\Delta\mathrm{S}$ variable is greatly overestimated towards the DPS-sensitive region by the on-shell and off-shell KaTie predictions, while the variable $\Delta\phi_{3\mathrm{j}}^{\min}$ is well described by the PYTHIA 8 samples and by the off-shell KaTie sample, which uses the `MRW-CT10nlo` TMD.

All models, except the off-shell KaTie models, fail to describe the shape of the $\Delta\mathrm{Y}$ observable. For the ϕ_{ij} observable, the models show the same tendencies as found in the previous two sections. The LO 2 → 2 PYTHIA 8 models show too large of a decorrelation, while predictions using higher order matrix element calculations perform better, as seen in the shape of the KaTie curves.

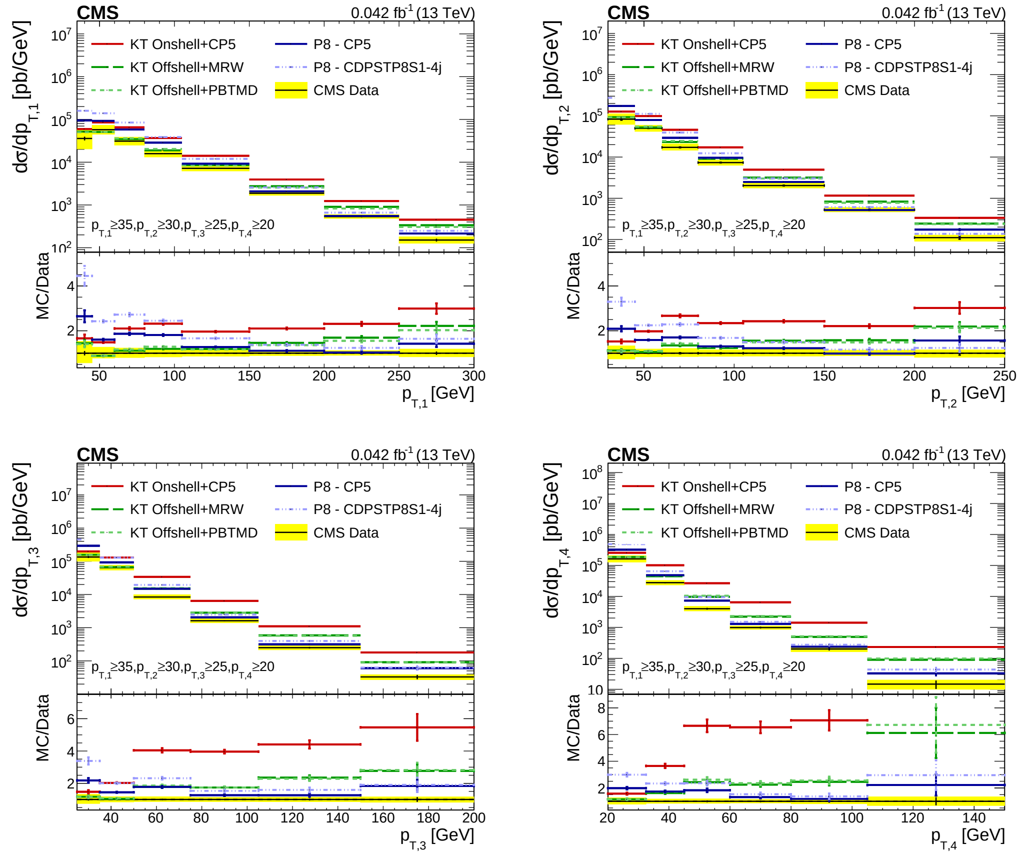


Figure 12.11: Comparison of the unfolded p_T spectra of the data with different SPS+DPS KaTie (KT) and PYTHIA 8 (P8) models, for the leading (top left), sub-leading (top right), third leading (bottom left), and fourth leading (bottom right) jet. The error bars represent the statistical uncertainties, and the yellow band indicates the total uncertainty on the measurement.

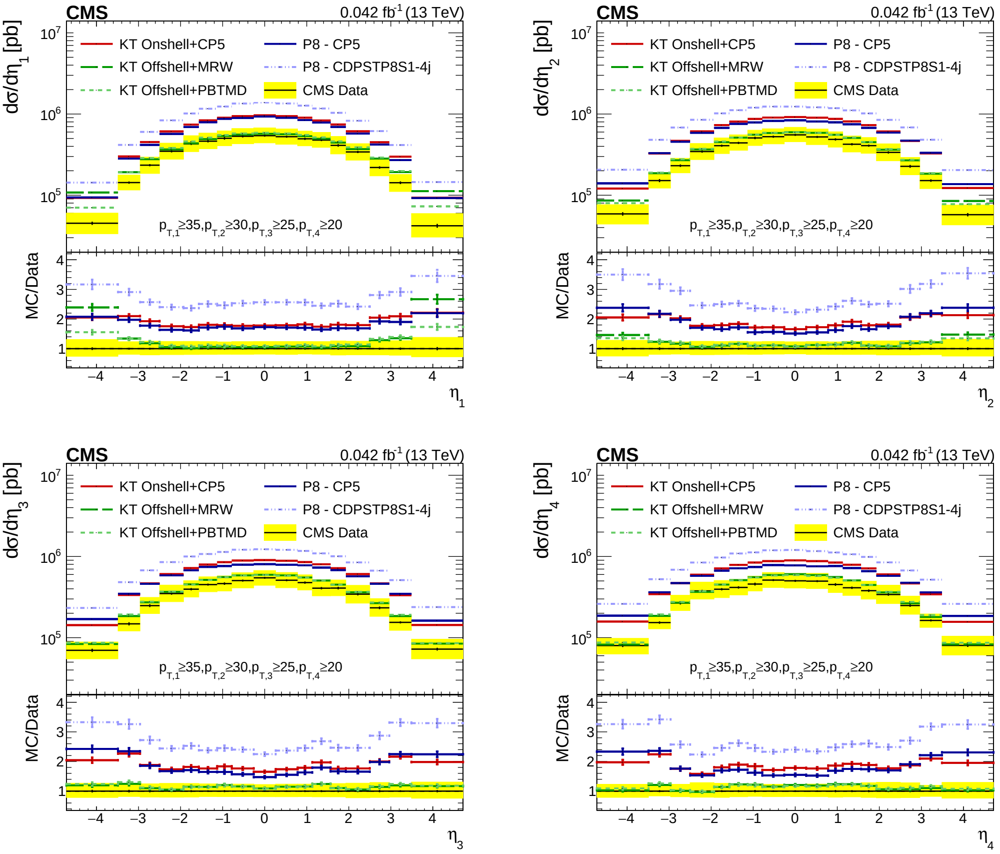


Figure 12.12: Comparison of the unfolded η spectra of the data with different SPS+DPS KaTie (KT) and pythia 8 (P8) models, for the leading (top left), sub-leading (top right), third leading (bottom left), and fourth leading (bottom right) jet. The error bars and bands are shown similarly to Fig. 12.11.

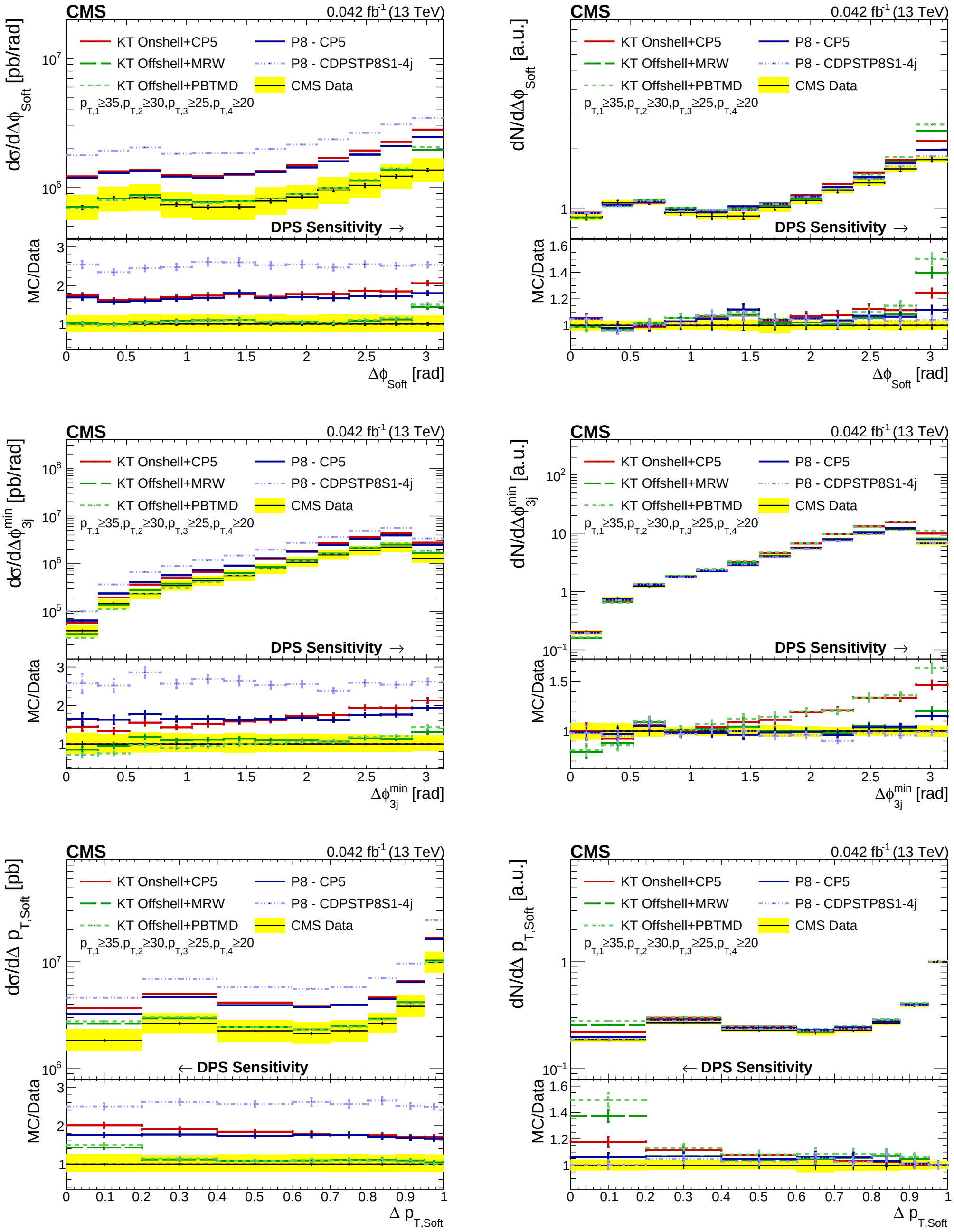


Figure 12.13: Comparison of the unfolded $\Delta\phi_{\mathrm{soft}}$, $\Delta\phi_{3\mathrm{j}}^{\mathrm{min}}$ and $\Delta p_{\mathrm{T,soft}}$ observables of the data with the SPS+DPS KaTie (KT) and pythia 8 (P8) models. The left-hand plots are in terms of the absolute cross section, while the right-hand plots are normalized to the region where reduced DPS-sensitivity is expected. The error bars and bands are shown similarly to Fig. 12.11.

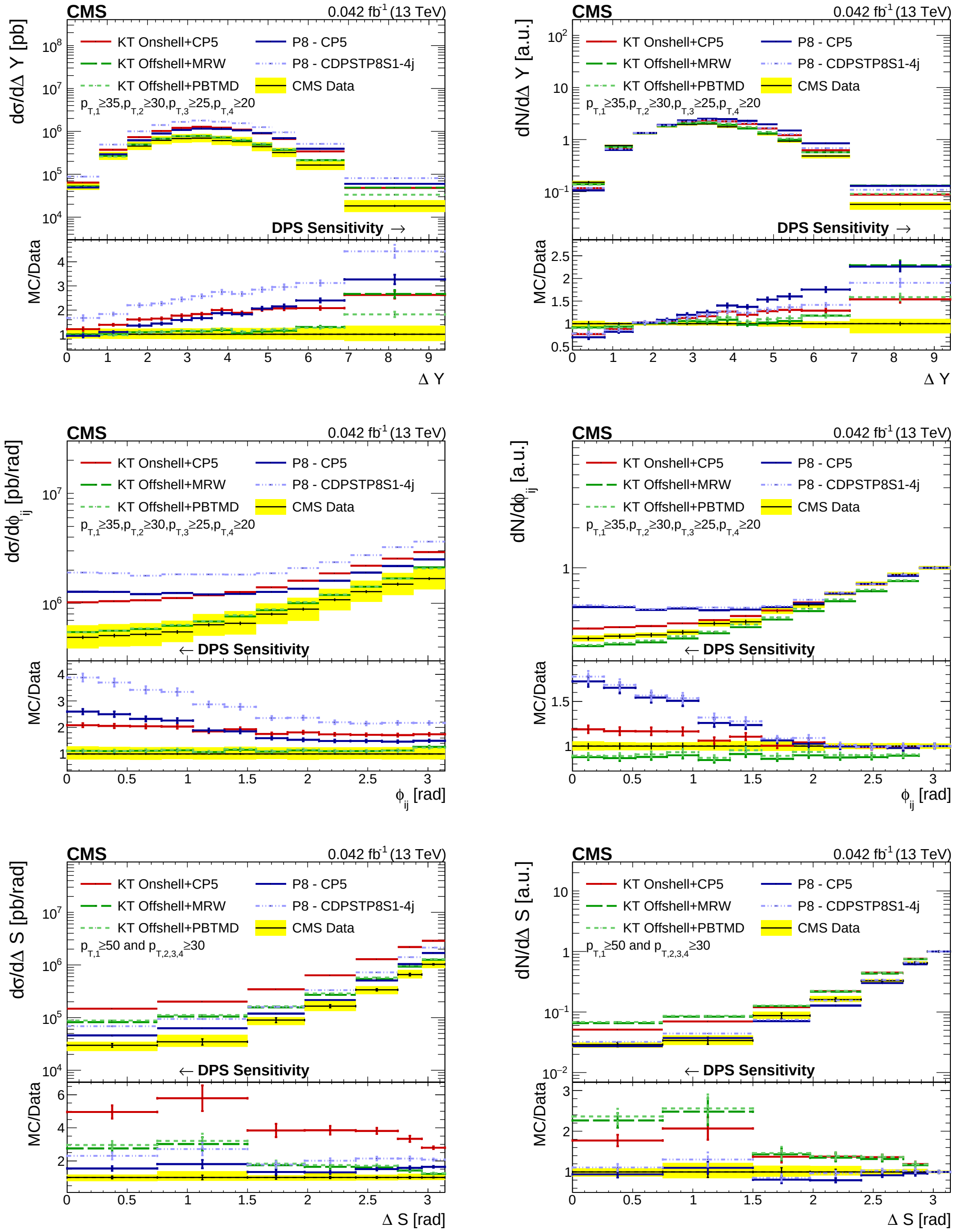


Figure 12.14: Comparison of the unfolded ΔY, ϕ_{ij} and ΔS observables of the data with the SPS+DPS KaTie (KT) and PYTHIA 8 (P8) models. The left-hand plots are in terms of the absolute cross section, while the right-hand plots are normalized to the region where reduced DPS-sensitivity is expected. The observable ΔS has been determined in *region II*, compared to *region I* for all other observables. The error bars and bands are shown similarly to Fig. 12.11.

12.5 Extraction of the Effective Cross Section

The DPS cross section is determined, and the effective cross section σ_{eff} is extracted through the use of the SPS MC models while following the two methods laid out in Chap. 8. The PYTHIA 8 and HERWIG samples using the `CUETP8M1` and the `CUETHS1` are omitted since the newer and improved `CP5` and `CH3` tunes are available. The extraction is also performed with the PYTHIA 8 sample using the `CDPSTP8S1-4j` tune, which already contains a DPS contribution. In this manner, the performance of the tune at a centre-of-mas energy of $\sqrt{s} = 13$ TeV is tested. The ΔS observable has proven to be the most robust of the DPS-sensitive observables since it is least affected by the type of parton shower. The difference between the ΔS observable and the other DPS-sensitive observables lies in its definition. All the information of the four leading jets enters the ΔS definition, which is not the case for all other DPS-sensitive observables. In the extraction procedure, the multijet KATIE models will not be considered; they greatly overshoot the DPS-sensitive slope of the ΔS observable which would result in non-convergent results.

12.5.1 Template Method

The pure DPS sample, which has been constructed from the data as described in Sec. 8.1.2, is used to obtain the ΔS distributions for pure DPS events, denoted as $\Delta\text{S}_{\text{DPS}}$. The $\Delta\text{S}_{\text{DPS}}$ distribution is corrected through the unfolding procedure in exactly the same manner as the other observables. The corrected $\Delta\text{S}_{\text{DPS}}$ distribution is shown in Fig. 12.15, along with the $\Delta\text{S}_{\text{DPS}}$ distributions of the DPS contribution from the PYTHIA 8 and on-shell KATIE models, both interfaced with PYTHIA 8 using the `CP5` tune. All distributions have been normalized to unity. The DPS data sample exhibits a larger decorrelation compared to the DPS MC samples. The differences can be attributed to the disparities between the p_{T} spectra of the data and the MC samples.

The template fitter object [176] takes the SPS MC distributions along with the DPS data sample as an input and determines the fraction of DPS events f_{DPS}. Through the use of eq. (8.5) and eq. (8.7), the DPS cross section and the effective cross section are extracted. The results for the considered models are given in Tab. 12.4 and are summarized for σ_{eff} in Fig. 12.16. The fitted distributions for the different models along with their statistical and systematical uncertainties are shown in Fig. 12.17 and in Fig. 12.18 for the PYTHIA 8 and HERWIG, and the multijet models, respectively.

The systematic uncertainties on the DPS fraction f_{DPS} have been determined through the use of the template fitter. The correlations of the systematic uncertainties have been taken into account. The ΔS and $\Delta\text{S}_{\text{DPS}}$ distributions are simultaneously varied for each systematic uncertainty individually. Next, the varied fractions are compared to their nominal value. Afterwards, the difference is taken as the systematic uncertainty due to the considered source. Finally, the total systematic uncertainty is calculated by taking the squared sum of all systematic uncertainties. The systematic uncertainties of the DPS cross section $\sigma^{\text{DPS}}_{\text{A,B}}$ and the effective cross section σ_{eff} have been determined in the same manner. Each varied DPS fraction enters eq. (8.7) to obtain the varied DPS cross section, which in turn is used as input in eq. (8.5), along with all the other varied objects. The remainder of the steps in the calculation include comparing the varied objects to their nominal values and taking the squared sum of all differences to obtain the total systematical uncertainty.

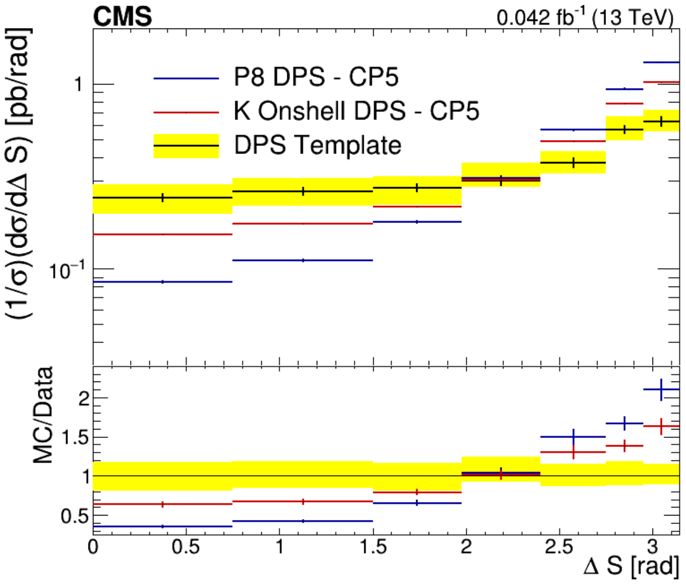


Figure 12.15: Comparison of the unfolded ΔS_{DPS} observable of the pure DPS data sample with the DPS MC models. The distribution has been determined in the phase space domain *region II*, while the distributions are normalized to unity. The error bars represent the statistical uncertainties, and the yellow band indicates the total uncertainty on the measurement.

The fraction of DPS events ranges from 0.0241 up to 0.1036, ignoring the value of -0.01294 obtained with the PYTHIA 8 sample using the `CDPSTP8S1-4j` tune. When examining the values for the DPS fraction f_{DPS}, the DPS cross section $\sigma^{\mathrm{DPS}}_{\mathrm{A,B}}$ and the effective cross section σ_{eff}, a clear distinction between four groups of models can be made.

- A first group consists of the PYTHIA 8 model with the dipole-antenna showering and the HERWIG 7 model using the `SoftTune`. Both models use an older PDF and tune, resulting in smaller DPS fractions.

- The second group of models are the PYTHIA 8 and the HERWIG 7 samples using the `CP5` and `CH3` tunes, respectively. The effective cross section σ_{eff} obtained with these models agrees within the total uncertainty with the four jet measurement performed by the CMS collaboration at a centre-of-mass energy of $\sqrt{s} = 7$ TeV where a value of $\sigma_{\mathrm{eff}} = 21.3^{+1.2}_{-1.6}$ mb was found [121].

- The two MADGRAPH 5 LO $2 \to 2, 3, 4$ samples form the third distinct group of models. The value of σ_{eff} shows agreement with the four jet measurements performed by the ATLAS collaboration at a centre-of-mass energy of $\sqrt{s} = 7$ TeV, where the value of $\sigma_{\mathrm{eff}} = 14.9^{+1.2}_{-1.0}(\mathrm{stat.})^{+5.1}_{-3.8}(\mathrm{sys.})$ mb was extracted [24].

- The fourth and last category consists of the NLO models for which the highest DPS fraction and lowest values of σ_{eff} are extracted.

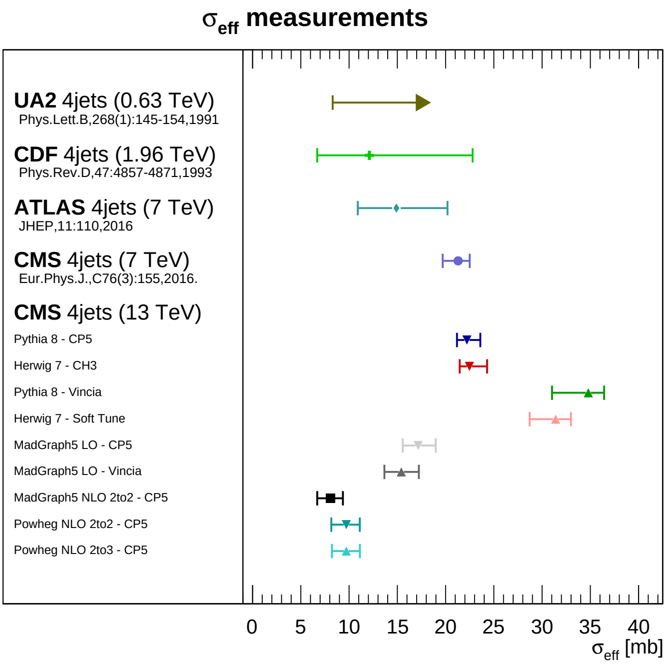


Figure 12.16: Comparison of all the extracted values of σ_{eff} for the template method for different models, along with the results from prior four jet measurements, performed at lower centre-of-mass energies. The values of the σ_{eff} of the older measurements are taken from [21], [22], [25] and [121] for the UA2, CDF, ATLAS and CMS collaborations, respectively.

All measurements agree with the lower bound of $\sigma_{\text{eff}} = 8.3$ mb, set by the measurement performed by the UA2 collaboration [21]. While all models, except the PYTHIA 8 model with the dipole-antenna showering provided by VINCIA and the HERWIG 7 model using the `SoftTune`, agree with the measurement performed by the CDF collaboration at a centre-of-mass energy of $\sqrt{s} = 1.96$ TeV, where a σ_{eff} equal to $12.1^{+10.7}_{-5.4}$ mb was found [22].

Two other DPS measurements have been performed at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. A value of σ_{eff} equal to $12.7^{+5.0}_{-2.9}$ mb has been extracted from the same-sign WW measurement [15]. while the J/Ψ pair production measurement found a $\sigma_{\text{eff}} = 7.3 \pm 0.5(\text{stat.}) \pm 1.0(\text{sys.})$ mb [20]. Although, completely different processes are measured in the two aforementioned studies, it is noteworthy that rather low values of σ_{eff} were found compared to measurements at lower centre-of-mass energies. The former measurement agrees within uncertainty with all predictions made by the NLO and MADGRAPH 5 LO models, except for σ_{eff} predicted by the MADGRAPH 5 NLO model, which is the only model to agree with the J/Ψ pair production measurement. The results are extracted from three completely different DPS channels, however, the parameter σ_{eff} is expected to be universal and process independent when considering the fully inclusive scenarios [108].

The interplay between the parton shower and the models using different matrix elements is apparent, just as it was for the DPS-sensitive variables. The PYTHIA 8 and HERWIG 7 models all use a 2 → 2 ME for which the two softest jets will always originate from the parton shower. These soft jets exhibit correlations similar to those of jets from DPS events, therefore, resulting in a significantly smaller discrepancy with the data compared to the other models, effectively yielding high values of the effective cross section σ_{eff}. The MADGRAPH 5 LO $2 \rightarrow 2, 3, 4$ samples experience the same effect as the PYTHIA 8 and HERWIG 7 models. A large fraction of events will be produced with a 2 → 2 ME. Due to the inclusion of higher order ME, the correlations between the softest jets will differ from those solely arising from jets coming from the parton showers. Lastly, the difference between the MADGRAPH 5 LO $2 \rightarrow 2, 3, 4$ and NLO samples is due to the inclusion of the NLO corrections. These three models allow for ME of higher order than the 2 → 2 ME at LO and include first order loop diagrams, which are absent from all other models.

The PYTHIA 8 model using the `CDPSTP8S1-4j` tune already contains a DPS contribution and uses a value of σ_{eff} equal to 21.3 mb [121]. In this analysis, an excess of DPS events is found, yielding a negative DPS fraction $f_{\text{DPS}} = -0.0129 \pm 0.0008 \,^{+0.0047}_{-0.0081}$. Through the combination of eq. (8.5) and eq. (8.7), along with a value of 21.3 mb for the σ_{eff} parameter, the corresponding DPS fraction $f^{\text{tune}}_{\text{DPS}}$ can be extracted. A value of $f^{\text{tune}}_{\text{DPS}} = 0.0393 \pm 0.0009 \,^{+0.0047}_{-0.0066}$ was found. The corrected fraction of DPS events needed by the model can now be defined as

$$f^{\text{true}}_{\text{DPS}} = f^{\text{tune}}_{\text{DPS}} + f_{\text{DPS}} \tag{12.5}$$

Resulting in a true DPS fraction of $f^{\text{true}}_{\text{DPS}} = 0.0263 \pm 0.001 \,^{+0.007}_{-0.010}$. The corrected effective cross section corresponding to the newly determined fraction of DPS events is equal to $31.76 \pm 1 \,^{+5}_{-19}$. This value is consistent with the results obtained from the PYTHIA 8 model with the dipole-antenna showering from VINCIA and the HERWIG 7 model using the `SoftTune`. To better fit the four jet measurement at 13 TeV the PYTHIA 8 sample using the `CDPSTP8S1-4j` would need a higher σ_{eff}.

Table 12.4: The values of the DPS fraction f_{DPS}, the DPS cross section $\sigma_{\mathrm{A,B}}^{\mathrm{DPS}}$ and the effective cross section σ_{eff} for the different LO and NLO models, along with their statistical and systematical uncertainties. All results have been obtained through the template method.

MC Model	Tune	$f_{\mathrm{DPS}} \pm$ (stat.) $\pm$ (syst.)	$\sigma_{\mathrm{A,B}}^{\mathrm{DPS}}$ $\pm$ (stat.) $\pm$ (syst.) (nb)	σ_{eff} $\pm$ (stat.) $\pm$ (.syst.) (mb)
pythia 8 and herwig Samples				
pythia 8	`CP5`	$0.0377 \pm 0.0008\ ^{+0.0045}_{-0.0068}$	$22.9 \pm 0.7\ ^{+5.7}_{-7.3}$	$22.2 \pm 0.7\ ^{+1.2}_{-0.8}$
pythia 8	`CDPSTP8S1-4j`	$-0.0129 \pm 0.0008\ ^{+0.0047}_{-0.0081}$	/	/
pythia 8+Vincia	Standard pythia 8.3	$0.0241 \pm 0.0007\ ^{+0.0041}_{-0.0068}$	$14.6 \pm 0.6\ ^{+4.4}_{-5.9}$	$34.8 \pm 1.3\ ^{+1.0}_{-3.5}$
herwig 7	`CH3`	$0.0372 \pm 0.0007\ ^{+0.0038}_{-0.0068}$	$22.6 \pm 0.7\ ^{+5.1}_{-7.3}$	$22.5 \pm 0.7\ ^{+1.7}_{-0.7}$
herwig 7	`SoftTune`	$0.0267 \pm 0.0007\ ^{+0.0042}_{-0.0071}$	$16.2 \pm 0.6\ ^{+4.6}_{-6.3}$	$31.4 \pm 1.1\ ^{+1.1}_{-2.4}$
Multijet Samples				
MadGraph 5 LO, pythia 8	`CP5`	$0.0488 \pm 0.0007\ ^{+0.0036}_{-0.0064}$	$29.6 \pm 0.8\ ^{+5.7}_{-8.2}$	$17.2 \pm 0.5\ ^{+1.8}_{-1.5}$
MadGraph 5 LO, pythia 8+Vincia	Standard pythia 8.3	$0.0543 \pm 0.0007\ ^{+0.0033}_{-0.0060}$	$33.0 \pm 0.8\ ^{+5.9}_{-8.6}$	$15.4 \pm 0.4\ ^{+1.8}_{-1.7}$
MadGraph 5 NLO, pythia 8	`CP5`	$0.1036 \pm 0.0006\ ^{+0.0026}_{-0.0050}$	$63 \pm 1\ ^{+9}_{-13}$	$8.1 \pm 0.2\ ^{+1.3}_{-1.4}$
powheg NLO 2 → 2, pythia 8	`CP5`	$0.0857 \pm 0.0004\ ^{+0.0030}_{-0.0052}$	$52 \pm 1\ ^{+8}_{-11}$	$9.8 \pm 0.2\ ^{+1.4}_{-1.5}$
powheg NLO 2 → 3, pythia 8	`CP5`	$0.0862 \pm 0.0007\ ^{+0.0030}_{-0.0058}$	$52 \pm 1\ ^{+8}_{-12}$	$9.7 \pm 0.2\ ^{+1.4}_{-1.5}$

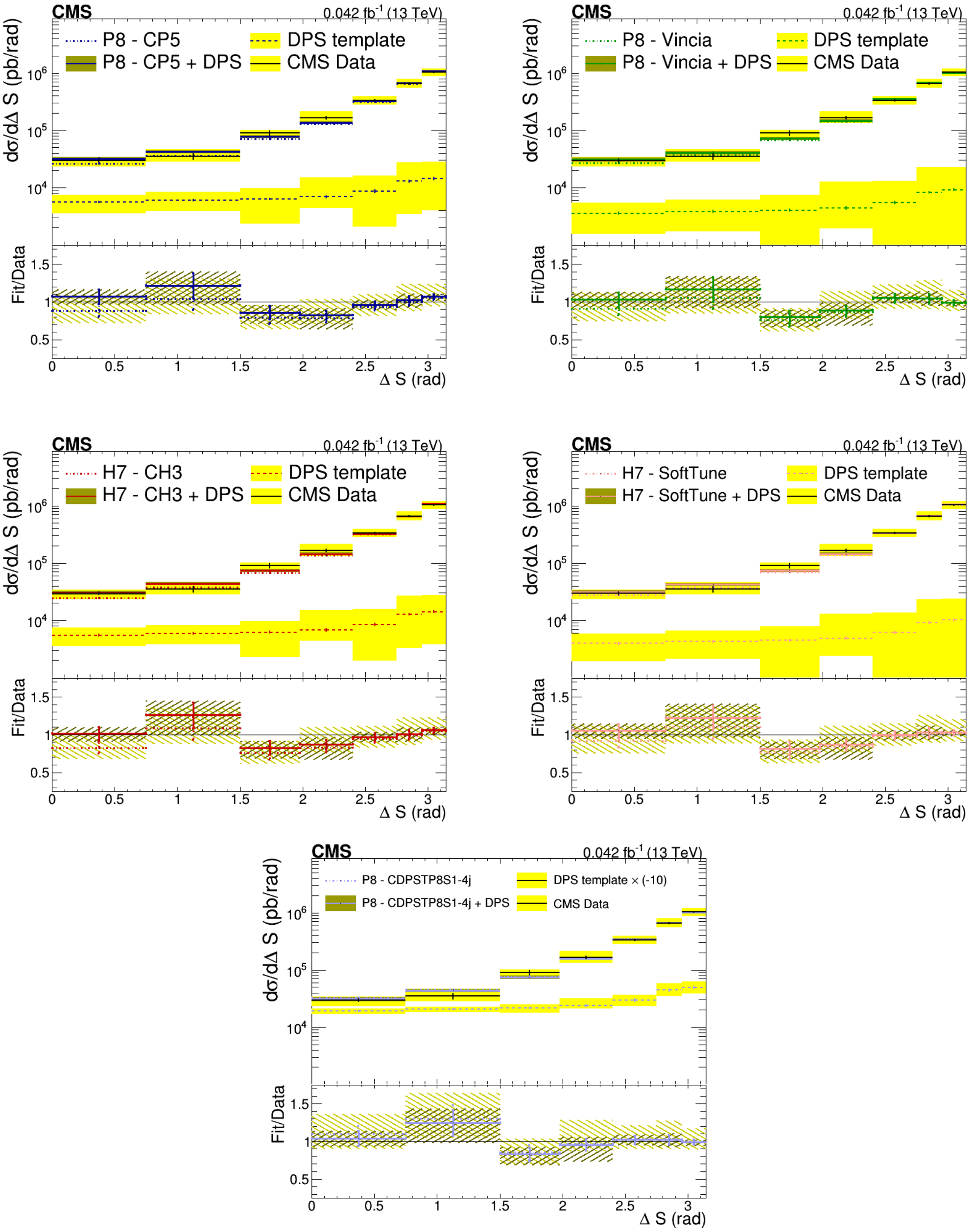


Figure 12.17: The results of the template fit for the different PYTHIA 8 (P8) and HERWIG 7 (H7) models. The yellow bands represent the total uncertainty of the distribution. The ratio of both the fitted MC model and the total fitted result relative to the data are show. As the ΔS_{DPS} carries a statistical and systematical uncertainty, so does the total fitted sample. The results for the PYTHIA 8 model using the `CDPSTP8S1-4j` tune has been scaled with a factor −10 since a negative DPS fraction f_{DPS} was found.

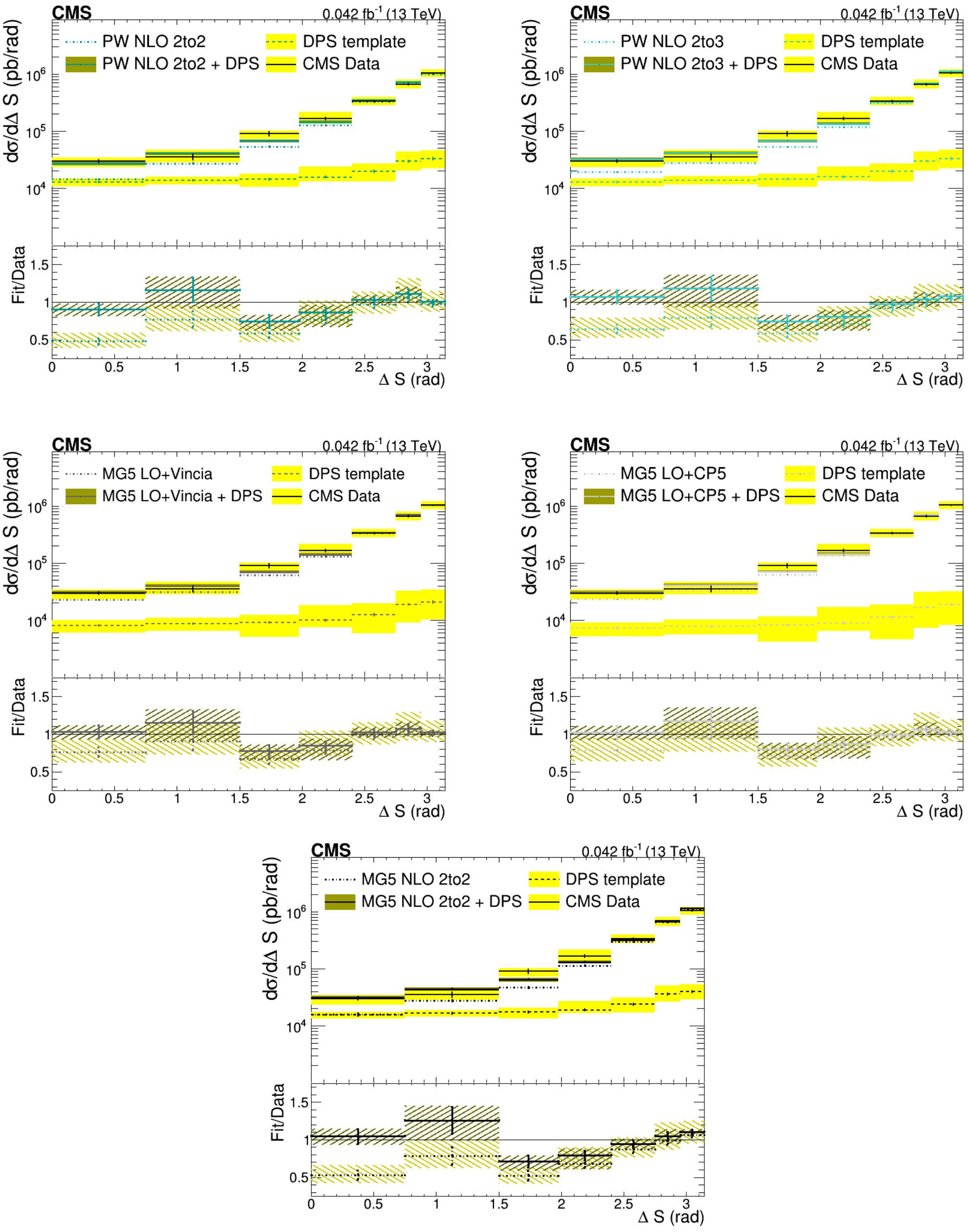


Figure 12.18: The results of the template fit for the different multijet models. All error bands and bars on both subplots for each model are similarly to Fig. 12.17.

12.5.2 Shape Comparison Method

The second method, detailed in Sec. 8.2, aims to exploit the bin-to-bin differences between the data distribution and the predictions from the MC models of the ΔS observable. The bin-normalized distributions are used since they allow for a direct shape comparison, while having smaller uncertainties compared to the distributions in terms of the absolute cross section. In Fig. 12.19, the bin-normalized distributions for the PYTHIA 8 and HERWIG (left), and multijet (right) models are bundled together along with the data for clarity's sake. The fraction of DPS events f_{DPS} can be estimated by considering the difference between the bin-normalized distributions as shown in eq. (8.8). Together with eq. (8.5), eq. (8.8), the DPS cross section $\sigma^{\mathrm{DPS}}_{\mathrm{A,B}}$ and the effective cross section σ_{eff} are determined. The results for all the considered models are given in Tab. 12.5.

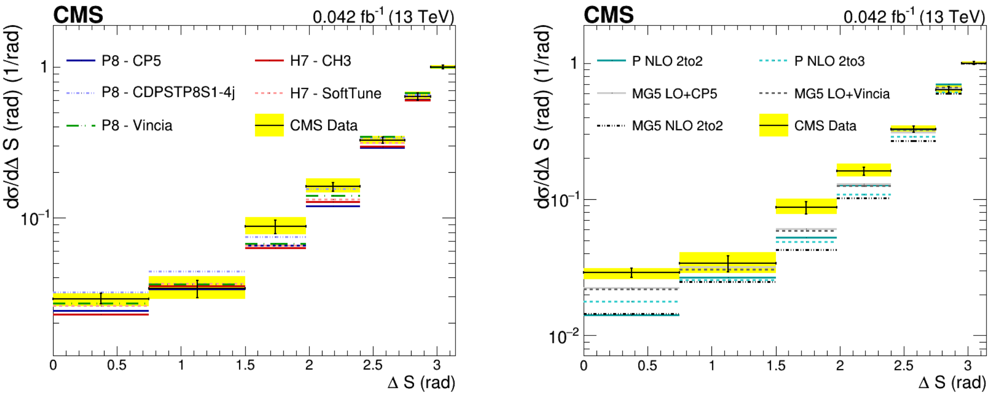


Figure 12.19: Comparison of the predictions of the different SPS MC models for the bin-normalized ΔS observable along with the distribution obtained from the data.

The systematic uncertainties on f_{DPS}, $\sigma^{DPS}_{\mathrm{A,B}}$ and σ_{eff} have been determined in a similar manner compared to the results of the template fit in the previous section. For each systematic uncertainty, all distributions are varied accordingly in order to take the correlations into account. The varied results are compared to their nominal results. The total uncertainty is the squared sum of all discrepancies between the varied and nominal values of the objects.

When examining the results in Tab. 12.5, it immediately stands out that the statistical and systematical errors are of the same order, which is due to the limited available statistics in the phase space domain *region II*. A second remarkable result is the sensitivity to small bin-to-bin differences. A first example of this sensitivity is shown by the predictions made by the PYTHIA 8 model with a dipole-antenna showering provided by VINCIA. Some of the bin values of the bin-normalized distributions of the SPS MC models have a larger value compared to those from the data plot. Therefore, the resulting DPS fraction is small and the uncertainties are larger than the values for f_{DPS}, $\sigma^{\mathrm{DPS}}_{\mathrm{A,B}}$ and σ_{eff}. A second and more subtle example is found when examining the POWHEG NLO 2 → 2 predictions. The second to last bin of the model overshoots the data, while all bins in the DPS-sensitive tail lie lower compared to the MadGraph 5 LO 2 → 2, 3, 4 models. Nonetheless, the predicted values of σ_{eff} of the POWHEG NLO 2 → 2 and the MadGraph 5 LO 2 → 2, 3, 4 models agree with each other but not with the other NLO predictions.

The results for multijet models, except the POWHEG NLO 2 → 2 model, agree with the results obtained from the template method. Taking the two observations into account, along with the fact that a DPS contribution can be present in the bin the distributions are normalized to, it is opted to consider the results of the template method as the main outcome of the whole extraction effort.

Increased statistics will most certainly benefit the stability of the shape comparison method. Another workaround would be to implement more asymmetrical cuts, as it is expected to boost the DPS signature. In the extraction of σ_{eff}, cuts of 50-40-30-20 GeV or of 35-30-25-20 GeV were not possible due to constraints imposed by the available triggers.

Table 12.5: The values of the DPS fraction f_{DPS}, the DPS cross section $\sigma^{\text{DPS}}_{\text{A,B}}$ and the effective cross section σ_{eff} for the different LO and NLO models, along with their statistical and systematical uncertainties. All results have been obtained through the ΔS shape comparison method.

Sample	Tune	$f_{\text{DPS}} \pm (\text{stat.})^{+\text{U.sys.}}_{\text{-L.sys.}}$	$\sigma^{\text{DPS}}_{\text{A,B}} \pm (\text{stat.})^{+\text{U.sys.}}_{\text{-L.sys.}}$ (nb)	$\sigma_{\text{eff}} \pm (\text{stat.})^{+\text{U.sys.}}_{\text{-L.sys.}}$ (mb)
PYTHIA 8 and HERWIG Samples				
PYTHIA 8	`CP5`	$0.086 \pm 0.019^{+0.018}_{-0.022}$	$53 \pm 12^{+12}_{-12}$	$9.7 \pm 2.2^{+3.6}_{-1.1}$
PYTHIA 8	`CDPSTP8S1`	$-0.016 \pm 0.020^{+0.023}_{-0.026}$	\	\
PYTHIA 8+Vincia	Standard PYTHIA 8.3	$0.008 \pm 0.021^{+0.016}_{-0.012}$	$5 \pm 13^{+16}_{-1}$	$(11 \pm 28^{+14}_{-6}) \times 10$
HERWIG 7	`SoftTune`	$0.047 \pm 0.020^{+0.019}_{-0.028}$	$29 \pm 12^{+12}_{-10}$	$117.7 \pm 7.6^{+7.5}_{-1.9}$
HERWIG 7	`CH3`	$0.080 \pm 0.020^{+0.019}_{-0.026}$	$49 \pm 12^{+12}_{-14}$	$10.5 \pm 2.6^{+4.2}_{-1.1}$
Multijet Samples				
MadGraph 5 LO, PYTHIA 8	`CP5`	$0.058 \pm 0.020^{+0.020}_{-0.027}$	$35 \pm 12^{+12}_{-9}$	$14.3 \pm 4.9^{+6.6}_{-1.5}$
MadGraph 5 LO, PYTHIA 8+Vincia	Standard PYTHIA 8.3	$0.054 \pm 0.020^{+0.020}_{-0.022}$	$33 \pm 12^{+12}_{-9}$	$15.4 \pm 5.7^{+6.9}_{-1.7}$
MadGraph 5 NLO, PYTHIA 8	`CP5`	$0.157 \pm 0.018^{+0.017}_{-0.020}$	$95 \pm 11^{+13}_{-12}$	$5.34 \pm 0.6^{+1.3}_{-0.6}$
POWHEG NLO 2 → 2, PYTHIA 8	`CP5`	$0.061 \pm 0.020^{+0.020}_{-0.020}$	$36 \pm 12^{+13}_{-9}$	$13.8 \pm 4.6^{+4.6}_{-1.7}$
POWHEG NLO 2 → 3, PYTHIA 8	`CP5`	$0.126 \pm 0.019^{+0.018}_{-0.022}$	$77 \pm 11^{+13}_{-12}$	$6.62 \pm 1.0^{+1.8}_{-0.7}$

Chapter 13

Summary

In this thesis, a measurement of the inclusive four jet cross section in proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV has been performed. The measurement aims to improve the understanding of the strong interaction, described by the theory of Quantum Chromodynamics, and is especially relevant from the perspective of double parton scattering (DPS), being the simplest form of multiple partonic interactions (MPI). The four leading jets are divided in a hard and soft jet pair, whose correlations are exploited in an effort to disentangle single parton scattering (SPS) events from DPS events through specifically chosen observables besides the transverse momentum and pseudorapidity spectra of the jets. The absolute cross sections have been measured for all observables. Additionally, the normalized differential cross sections for the DPS-sensitive observables have been determined in order to perform a more qualitative study of DPS. The differential cross sections have been normalized to the region of the observables that is expected to be insensitive to DPS.

The measurement was compared to three different sets of models, the first of which are the PYTHIA 8 and HERWIG models. The predicted cross sections of these models based on LO $2 \rightarrow 2$ matrix elements greatly overshoot the data, due to an excess of very forward/backward, low p_{T} jets. None of the individual models stands out, however a clear distinction between the models interfaced with a p_{T}-ordered shower and with an angular-ordered/dipole-antenna shower is observed in the $\Delta\phi_{3\mathrm{j}}^{\mathrm{min}}$ and ϕ_{ij} observables. The angular-ordered/dipole-antenna showered models predict a too large correlation in the $\Delta\phi_{3\mathrm{j}}^{\mathrm{min}}$ observable, while accurately describing the ϕ_{ij} observable. The opposite is true for the p_{T}-ordered shower. As the deviations from the data are much larger for models interfaced with a p_{T}-ordered shower compared to the models interfaced with an angular-ordered/dipole-antenna shower, one can conclude that the latter performs better throughout the phase space.

A second class of models consist of event generators using multi-leg and/or NLO matrix elements. The NLO and MadGraph 5 LO $2 \rightarrow 2, 3, 4$ samples generally provide a good description in terms of the absolute cross section, leaving room for a DPS contribution. However, differences between the models become apparent in the normalized distributions, in which the effect of different parton showers can again be observed. All NLO and MadGraph 5 LO $2 \rightarrow 2, 3, 4$ samples fail to give a good description of the shape of $\Delta\mathrm{Y}$, indicating room for improvement in the description of the pseudorapidity of the jets. A difference between the NLO models and the MadGraph 5 LO $2 \to 2, 3, 4$ models is seen in the normalized distributions. The NLO samples show a need for DPS events, whereas the correlations between the jets in the MadGraph 5 LO $2 \to 2, 3, 4$ samples seem to simulate similar correlations as expected from DPS events. The off-shell KaTie models perform similarly in terms of the absolute cross section and in the normalized distributions, giving a good description of the data throughout the phase space, except for the p_{T} spectra, which might be related to the use of a single $2 \to 4$ matrix element. The off-shell KaTie models looks promising and are expected to give an even better description of the p_{T} spectra, as soon as it is possible to mix matrix elements with a different number of final state partons using the KaTie MC event generator.

The third and last group of models are the models which have an explicit DPS contribution. From the normalized distributions one can conclude that the pythia 8 models perform fairly well throughout the phase space, except for the ϕ_{ij} observable. While the total KaTie samples, generated with a σ_{eff} equal to 21.3 mb^{-1} [121], overshoot the data in the DPS-sensitive region, indicating that the value used for σ_{eff} is too small. This might also be a consequence of using a single single $2 \to 4$ matrix element in the SPS sample.

The DPS pocket formula allows for the extraction of the effective cross section σ_{eff}, however, a workable form of the formula had to be derived first. In this regard, the two independent processes in a DPS event A and B have been defined as inclusive single jet processes and the 4-jet efficiency $\epsilon_{4\mathrm{j}}$ had to be incorporated in the formula, resulting in eq. (8.5). The $\Delta\mathrm{S}$ observable was most robust against parton shower effects and is, therefore, used in the extraction. Two independent methods have been implemented: the template method and the shape comparison method. Values of the effective cross section σ_{eff} have been extracted for all SPS models with both methods, except for the KaTie models. The shape comparison method yielded a much larger statistical uncertainty due to the limited available statistics compared to the template method. Additionally, a high sensitivity to small bin-to-bin differences was found, even in the region where only a small DPS contribution is expected. Therefore, the results of the template method are considered as the main outcome of the extraction effort. A clear model dependence is observed. Models where more jets originate from the parton shower tend to yield a larger σ_{eff}. A large effect due to the inclusion of NLO calculations is also observed, as the NLO models need the largest DPS contribution out of all models studied in this analysis.

Although, a wide range of values of the effective cross section σ_{eff} have been measured, it is notable that the predictions from the MadGraph 5 LO and powheg NLO samples agree within the uncertainty with the σ_{eff} extracted from the same-sign WW and the J/Ψ pair production DPS measurements, which have been performed at a centre-of-mass energy of $\sqrt{s} = 13$ TeV as well. The predictions from the MadGraph 5 NLO sample show agreement within the uncertainty with the J/Ψ pair production measurement. All measurements performed at a centre-of-mass energy of $\sqrt{s} = 13$ TeV, including this one, point to a value of σ_{eff} at the lower end of the spectrum since the DPS pocket formula is expected to be universal and process independent in the inclusive picture.

Overall no single model stands out in the description of DPS as many aspects of proton-proton collisions have subtle effects on the predictions. The effect of the parton shower is evident since DPS correlations can be simulated depending on the type of parton shower. The mixing of matrix elements or NLO calculations yield an improved description compared to LO $2 \to 2$ models, but can overestimate and underestimate the correlations in the data, respectively. The effect of the inclusion of off-shell initial states in the matrix elements looks promising, but leaves room for improvement as the mixing of different matrix element is not yet possible in such models.

Appendix A

Optimisation of MLM matching scale value

The MadGraph 5 LO samples that generates parton-level QCD processes with up to four partons in the final state, by combining 2 → 2, 2 → 3, 2 → 4 matrix elements, use a MLM matching scheme to take double counting effects into account. Such matching algorithms require to set a proper cutoff value (the matching scale) that should distinguish between jets from the ME and from the PS. However, that parameter does not have any physical meaning and its choice should not influence the final state observables in a significant way. A common way to check this is by analysing the differential jet rate (DJR) distributions. These show the transition from the region below the matching scale to the region above. The distributions of DJR have to be independent from the chosen cutoff, and the transition at the cutoff has to be as smooth as possible. More precisely the DJR plots are obtained by getting the distribution of the variable $\log(\sqrt{d_{ij}})$ where d_{ij} is the scale in a parton-level event where j jets are clustered into $i = j - 1$ jets using the k_{T}-algorithm. It thus measures the transition scale from (n+1)-jet to n-jet events and approximately reproduces the merging scale phase space.

The produced sample for this analysis uses the MLM scheme with k_{T} jet matching that has the following two main parameters that determine the performance of the matching algorithm:

- xqcut: minimum jet measure ($\mathrm{p_T}/k_{\mathrm{T}}$) for QCD partons used during the k_{T} jet matching. This value should be related to the hard scale (in our case H_{T}, required to be above 50 GeV) in the process, and set to 1/6th to 1/3th of the hard scale. This parameter is set in MadGraph 5.

- Qcut: This is the jet measure cutoff used by pythia 8 when it executes the matching algorithm after the parton showers have been created, and is set to max(xqcut+5, xqcut*1.2).

In our study, we will then vary these two parameters and see how the DJR plots behave. The values for which the plots are smooth and an efficient matching performance is obtained are chosen for the final sample production. Note that for this analysis the MPI and hadronization steps in pythia 8 have to be turned off in order to see the transition regions. As we are interested in the performance of combining $2 \to 2$, $2 \to 3$, $2 \to 4$ ME events we will focus on the DJR distributions of d_{23} and d_{34}, which represent the transition scales to go from 3 to 2 jets, and 4 to 3 jets, respectively.

The following sets of parameters have been tested:

- xqcut = 10 GeV Qcut = 14 GeV
- xqcut = 12 GeV Qcut = 20 GeV
- xqcut = 14 GeV Qcut = 25 GeV
- xqcut = 18 GeV Qcut = 32 GeV
- xqcut = 36 GeV Qcut = 75 GeV

In figure A.1, we can see the results for the d_{23} (left) and d_{34} (right) distributions for the first three choices of the xqcut parameter (10, 12, 14 GeV in first, second, and third row). In figure A.2 we can see the remaining choices of xqcut, 18 (first row) and 36 (second row) GeV, for d_{23} (left) and d_{34} (right). The black curves in the figures contain all events in the sample, while the other curves only contain $2 \to 2$ (red), $2 \to 3$ (blue), and $2 \to 4$ (green) ME events.

We clearly see that for the d_{23} differential jet rate plots the total event sample (black curve) is mostly populated by $2 \to 2$ events (red) below a certain scale, and populated by $2 \to 3$ events (blue) above. The region of this transition is close to the value of the corresponding Qcut cutoff used in PYTHIA 8 during the matching (e.g. for Qcut = 14 GeV we have $\log(\sqrt{(14)^2}) = 1.15$). A similar behaviour is seen for the d_{34} differential jet rate plots, where the event sample is mostly populated by $2 \to 2$ and $2 \to 3$ events below the matching scale, and populated by $2 \to 4$ events (green) above.

It is also clear that the d_{23} DJR distributions show a discrepancy around this matching scale for too low values of a chosen xqcut and Qcut value, which implies that the matching is not optimal. For a xqcut value of 14 GeV the d_{23} distribution starts to improve, but a visible change in slope around the matching scale is still present. For a choice of xqcut = 18 GeV (first row in figure A.2), a more smooth DJR distribution is obtained for both d_{23} and d_{34} and, therefore, this value has been chosen to produce the final MadGraph 5 sample. Finally, we also had a look at the DJR distributions when a deliberately high value of xqcut and Qcut were set, i.e. 36 and 75 GeV, respectively. This is shown in the second row of figure A.2 and also leads, maybe surprisingly, to continuous distributions. However, the number of events surviving the matching algorithm is much smaller, and dominated by $2 \to 2$ events. This smaller matching efficiency is due to the too high kinematic cuts.

The choice of the matching scale value does not have a physical meaning and should not influence the final state observable in a significant way, but it is clear from the DJR distributions that when the value of xqcut and Qcut increases the relative contribution of lower jet multiplicity events ($2 \to 2$ and $2 \to 3$) increases. This will lead to a higher contribution of jets from parton showers, and that can influence certain observables in our 4-jet study. This is illustrated in figure A.3, which shows several observables used in the analysis. Predictions of a KaTie $2 \to 2$ and $2 \to 4$ sample are compared with predictions of MadGraph 5 $2 \to 2, 3, 4$ samples with different choices of matching scale parameters.

As expected, the ϕ_{ij} variable (especially sensitive to parton shower effects) shows a dependence on the matching scale value. With increasing values of the cutoff parameters, the contribution of jets from parton showers increases, leading to a change in the tail of the distribution. In such way we see that for a rather low matching scale value (xqcut = 10) the behaviour of the mixed Madgraph $2 \to 2, 3, 4$ sample lies more close to a pure $2 \to 4$ sample from KaTie, while with the chosen higher cut (xqcut = 18) the behaviour comes somewhat closer to a pure $2 \to 2$ sample.

The $\Delta\phi_{\mathrm{soft}}$ variable that is in particular sensitive to DPS contributions shows a rather stable behaviour with respect to the matching scale value. So the choice of this cutoff parameter does not have an impact on our extractions of a potential DPS contribution. Finally, we also show the leading jet $p_{\mathrm{T,1st}}$, and fourth jet $p_{\mathrm{T,4th}}$ spectra in the bottom row of figure A.3. While there is a small dependence on the choice of xqcut at low leading jet $p_{\mathrm{T,1st}}$, the fourth jet $p_{\mathrm{T,4th}}$ spectrum is stable with respect to the different values of xqcut and Qcut.

The following references were used to write down the content in this section: the MadGraph 5 online manual, the pythia 8 online manual and [84].

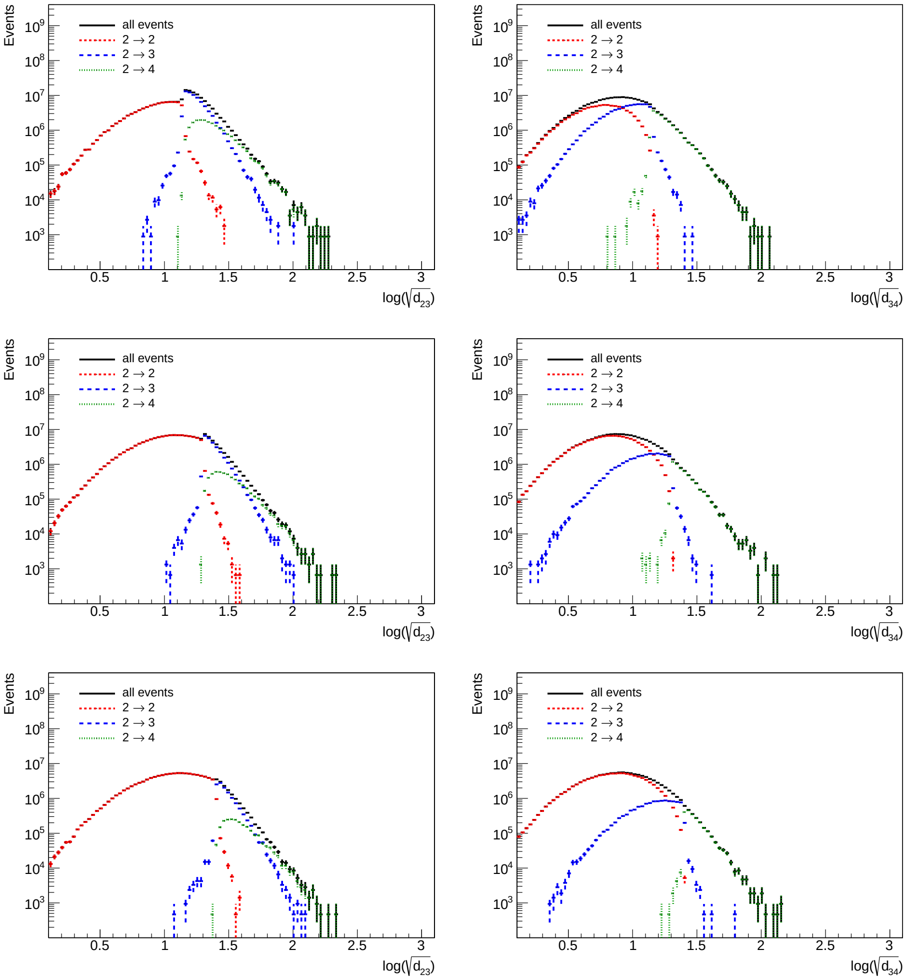


Figure A.1: DJR distributions for d_{23} (left) and d_{34} (right) for three choices of xqcut $=$ 10, 12, and 14 GeV presented in the first, second, and third row, respectively. The black curve contains all events in the sample, while the other curves only contain $2 \rightarrow 2$ (red), $2 \rightarrow 3$ (blue), and $2 \rightarrow 4$ (green) ME events.

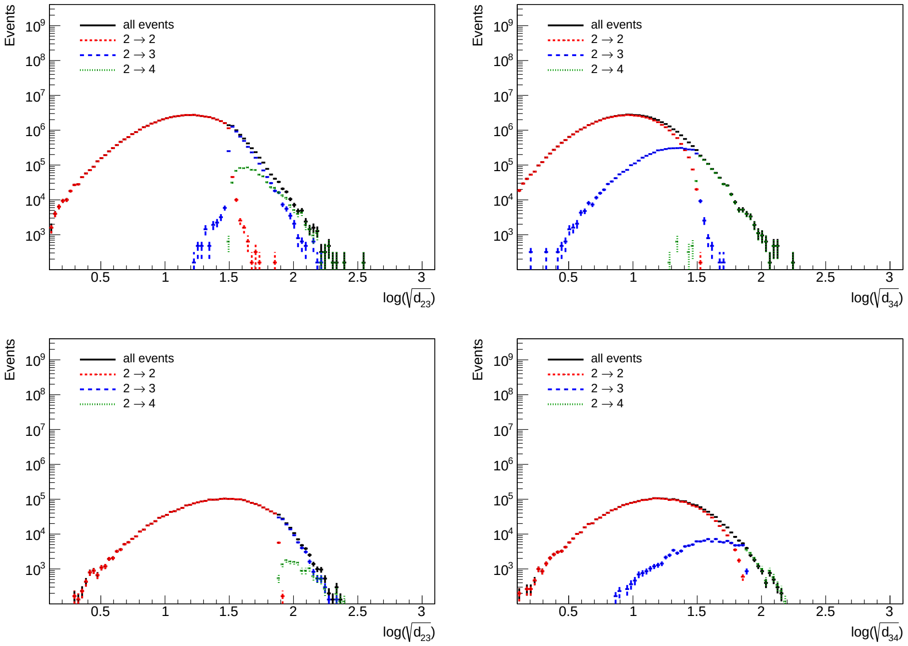


Figure A.2: DJR distributions for d_{23} (left) and d_{34} (right) for two choices of xqcut $= 18$ and 36 GeV presented in the first and second row, respectively. The black curve contains all events in the sample, while the other curves only contain 2 → 2 (red), 2 → 3 (blue), and 2 → 4 (green) ME events

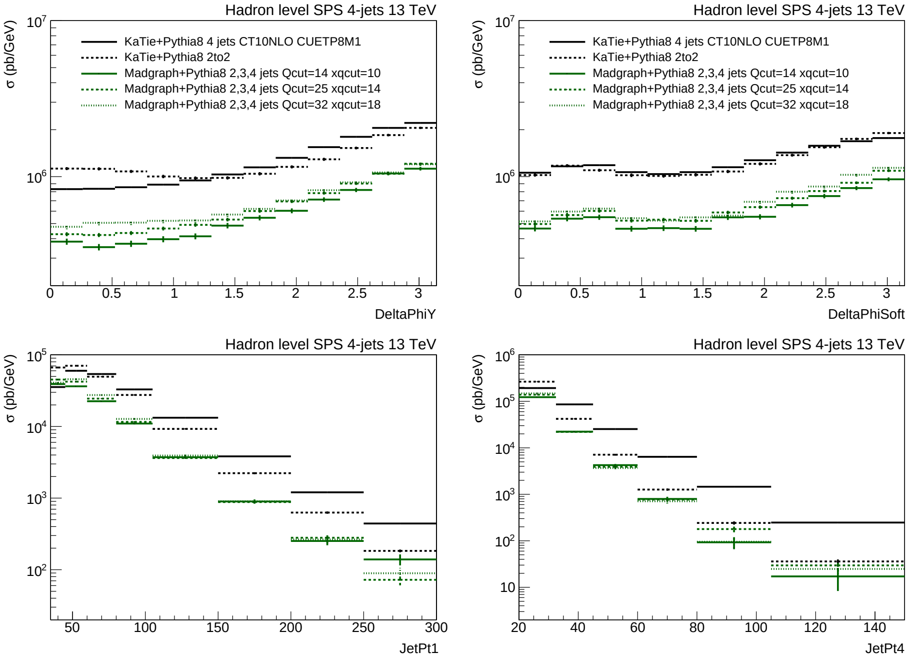


Figure A.3: The ϕ_{ij} (DeltaPhiY, top left), $\Delta\phi_{\mathrm{soft}}$ (DeltaPhiSoft, top right), leading jet $p_{\mathrm{T,1st}}$ (JetPt1, bottom left), and last leading jet $p_{\mathrm{T,4th}}$ (JetPt4, bottom right) distributions for different predictions of KaTie (with either $2 \to 2$ or $2 \to 4$ ME events) and MadGraph 5 $2 \to 2, 3, 4$ with different choices of matching scale.

Appendix B

KaTie off-shell DPS contributions at hadron level

In this section, we will briefly describe how we calculated the KaTie off-shell `MRW-CT10nlo` and `PB-NLO-HERAI+II-2018-set2` curves where both explicit SPS and DPS contributions are added to compare with unfolded data. The LHE files containing SPS predictions, where 2 → 4 matrix element calculations are used, obtained by KaTie can be fed into the Cascade MC to add initial state radiation, final state radiation and finally hadronization. The LHE files from KaTie containing DPS predictions, where two times 2 → 2 matrix element calculations are used, can unfortunately not be read out by Cascade with the current available program version. The main problem that arises here is that the color reconnection information is different and without any adjustment to treat such DPS events the existing Cascade MC framework is not able to process this information and will throw fatal event reconstruction errors. To calculate off-shell DPS contributions at hadron level we thus have to follow a different approach: we will use KaTie on-shell SPS and DPS predictions showered and hadronized with PYTHIA 8.240 to calculate the non-perturbative corrections, which are then applied to parton level off-shell KaTie DPS predictions. More precisely we will apply this to the relative DPS contributions, i.e. the ratio (SPS+DPS)/SPS. We will perform the following steps for this:

- Generate KaTie on-shell SPS and DPS predictions with `NNPDF31 nnlo` PDF and `CP5` tune at parton level, and gradually include initial state radiation, final state radiation, soft MPI and hadronization effects.
- Generate KaTie off-shell SPS and DPS predictions at parton level with `MRW-CT10nlo` and `PB-NLO-HERAI+II-2018-set2` TMD.
- Compare all these samples using the relative DPS contribution: $\mathrm{R} = (\mathrm{SPS}+\mathrm{DPS})/\mathrm{SPS}$.
- Calculate the non-perturbative corrections by taking the ratio: $\mathrm{C}_{\mathrm{NP}} = \mathrm{R}_{\mathrm{had}}^{\mathrm{on-shell}}/\mathrm{R}_{\mathrm{ISR}}^{\mathrm{on-shell}}$, using the on-shell KaTie samples, where the showering is handled by PYTHIA 8.
- Correct the relative DPS contribution of the parton level off-shell KaTie predictions $\mathrm{R}_{\mathrm{parton}}^{\mathrm{off-shell}}$ with the non-perturbative corrections: $\mathrm{R}_{\mathrm{had}}^{\mathrm{off-shell}} = \mathrm{C}_{\mathrm{NP}} \times \mathrm{R}_{\mathrm{parton}}^{\mathrm{off-shell}}$.

- Multiply the hadron level SPS KaTie off-shell prediction with $R_{\mathrm{had}}^{\mathrm{off-shell}}$ to get a curve that represents the off-shell SPS+DPS predictions at hadron level.

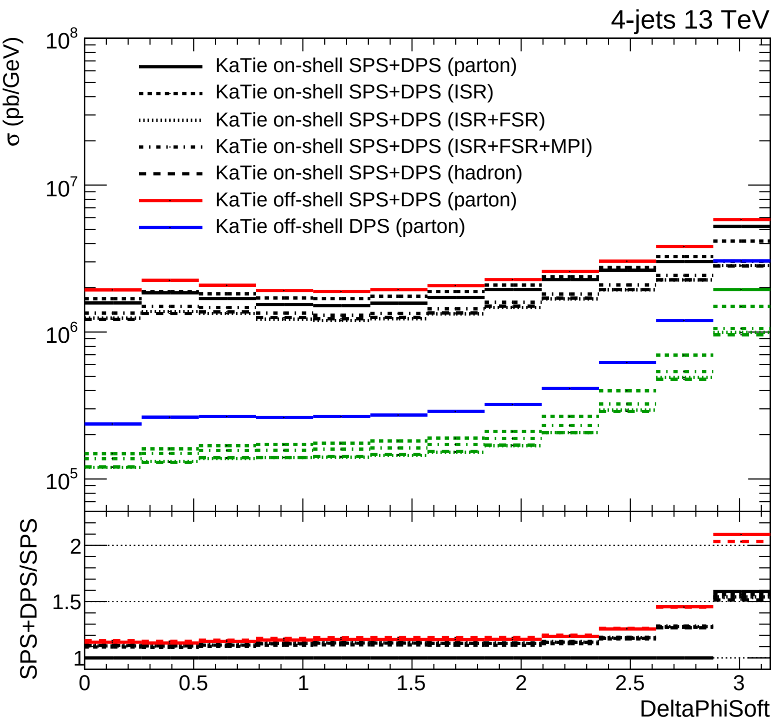


Figure B.1: Comparison of different on-shell KaTie predictions for the $\Delta\phi_{\mathrm{soft}}$ (DeltaPhiSoft) observable, and the parton level off-shell sample (for `MRW-CT10nlo` TMD). For the on-shell samples: black curves represent SPS+DPS predictions, and green curves DPS only predictions. The bottom panel presents the ratio (SPS+DPS)/SPS used to calculate the non-perturbative corrections.

In figure B.1, we can see the comparison of all the on-shell samples and the parton level off-shell sample with `MRW-CT10nlo` TMD for the observable $\Delta\phi_{\mathrm{soft}}$ (the azimuthal angle difference between the two soft jets) taken as example. The black curves represent the on-shell SPS+DPS predictions, and the green curves show the on-shell DPS only predictions. The solid lines are parton level samples, while the other line styles each present a different configuration of ISR, FSR, MPI and hadronization steps included. The parton level off-shell samples are presented by the red (SPS+DPS) and blue (DPS only) curves. The bottom panel shows the ratio R = (SPS+DPS)/SPS used to calculate the non-perturbative corrections.

It is directly clear that the DPS contribution of the on-shell parton level KaTie sample is concentrated in the last bin at pi, consistent with what we can expect from the kinematic limitations, whereas the DPS contributions becomes continuous when initial state radiation is included. Adding final state radiation, MPI and hadronization does not change the relative DPS contribution significantly for this observable. It is exactly this small difference that will be taken as the non-perturbative corrections: $C_{NP} = R_{had}^{on-shell}/R_{ISR}^{on-shell}$.

It is also interesting to see that the relative DPS contribution of the off-shell parton level sample ($R_{parton}^{off-shell}$, red solid curve in the bottom panel of figure B.1) is already continuous even without the presence of any initial state radiation. This is due to the initial k_T present in the used TMD. The red dashed curve shown in the bottom panel of figure B.1 is then the result of applying the derived non-perturbative corrections to the relative DPS contribution of the off-shell parton level samples ($R_{had}^{off-shell} = C_{NP} \times R_{parton}^{off-shell}$).

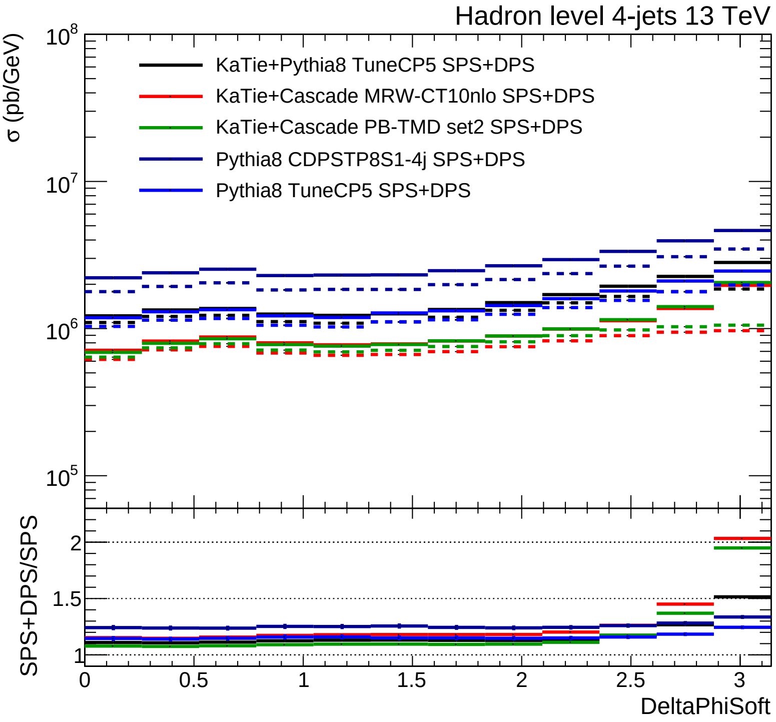


Figure B.2: Hadron level predictions of KaTie on-shell and off-shell samples, compared to standalone pythia 8 predictions for the $\Delta\phi_{soft}$ (DeltaPhiSoft) observable. The solid line curves present the total cross sections (SPS+DPS) and the dashed lines present the SPS only predictions. The off-shell KaTie SPS+DPS curves are obtained with the method described in the text. The bottom panel shows the relative DPS contribution: (SPS+DPS)/SPS.

The SPS off-shell KaTie samples generated at hadron level are then multiplied with this relative DPS contribution, $R_{had}^{off-shell}$, to construct a total SPS+DPS curve that should approximate the hadron level effects and can be compared with unfolded data. The final results for the off-shell KaTie samples with both the `MRW-CT10nlo` and `PB-NLO-HERAI+II-2018-set2` TMD are shown in figure B.2 and compared to the on-shell KaTie sample and two standalone pythia 8 samples interfaced with the `CDPSTP8S1-4j` and `CP5` tunes.

The above explained procedure is finally repeated for every observable to be compared with data. One effect that we can not take into account is the potential extra smearing, or de-correlation, due to adding initial state radiation on top of the parton level off-shell (SPS+DPS)/SPS curve. This could further smear out the relative DPS contribution and e.g. decrease the DPS fraction for the $\Delta\phi_{\text{soft}}$ observable near pi.

Appendix C

Additional JEC Plots

In Sec. 6.2, the L2L3 MC-Truth and L2L3 residual corrections have been derived. For clarity's sake, many of the control and closure plots were moved to this section.

All closure plots for each bin of the reconstructed p_{T} at particle level (p_{T}^{ptcl}) in function of the reconstructed pseudorapidity η^{Reco} are shown in appendix C.1 for the L2L3 MC-Truth corrections.

The control distributions of the variables A and B, defined in eq. (6.12) and eq. (6.14), respectively, are given in appendix C.2. In appendix C.3, the double ratios of the relative responses defined in eq. (6.11) and eq. (6.13) for the PYTHIA 8 sample relative to the data for both methods. The linear logarithmic fit, used to extract the p_{T} dependence of the double ratio, is also drawn on the plots. Before the corrections could be calculated, the radiation correction factor $k_{\mathrm{FSR}}(|\eta_{\mathrm{probe}}|)$ needed to be determined. The first step was to calculate $k_{\mathrm{FSR}}(\alpha_{\max}, |\eta_{\mathrm{probe}}|)$ for each bin in p_{T}. The results are shown in Fig. C.11 and Fig. C.12 in appendix C.4 for the p_{T}-balance and MPF method, respectively. Next, the weighted mean over all p_{T} bins in each $|\eta_{\mathrm{probe}}|$ and $\alpha_{\max}$ bin was calculated and fitted with a constant and linear fit for the interval $\alpha_{\max} \in [0.11, 0.32]$. Both fits are presented in Fig. C.14 and Fig. C.14, where it becomes clear why the fits have been restricted to the interval determined in Fig. 6.6. On the one hand, the two lowest bins vary wildly due to the limited amount of statistics. On the other hand, the three last bins were omitted from the fit as $\langle (p_{\mathrm{R,1st}}^{\mathrm{gen}} - p_{\mathrm{R,2nd}}^{\mathrm{gen}} / p_{\mathrm{T,Ave}}) \rangle / p_{\mathrm{T,Ave}}$ was found not to be linear in $\alpha_{\max}$.

C.1 L2L3 MC-Truth Closure Plots

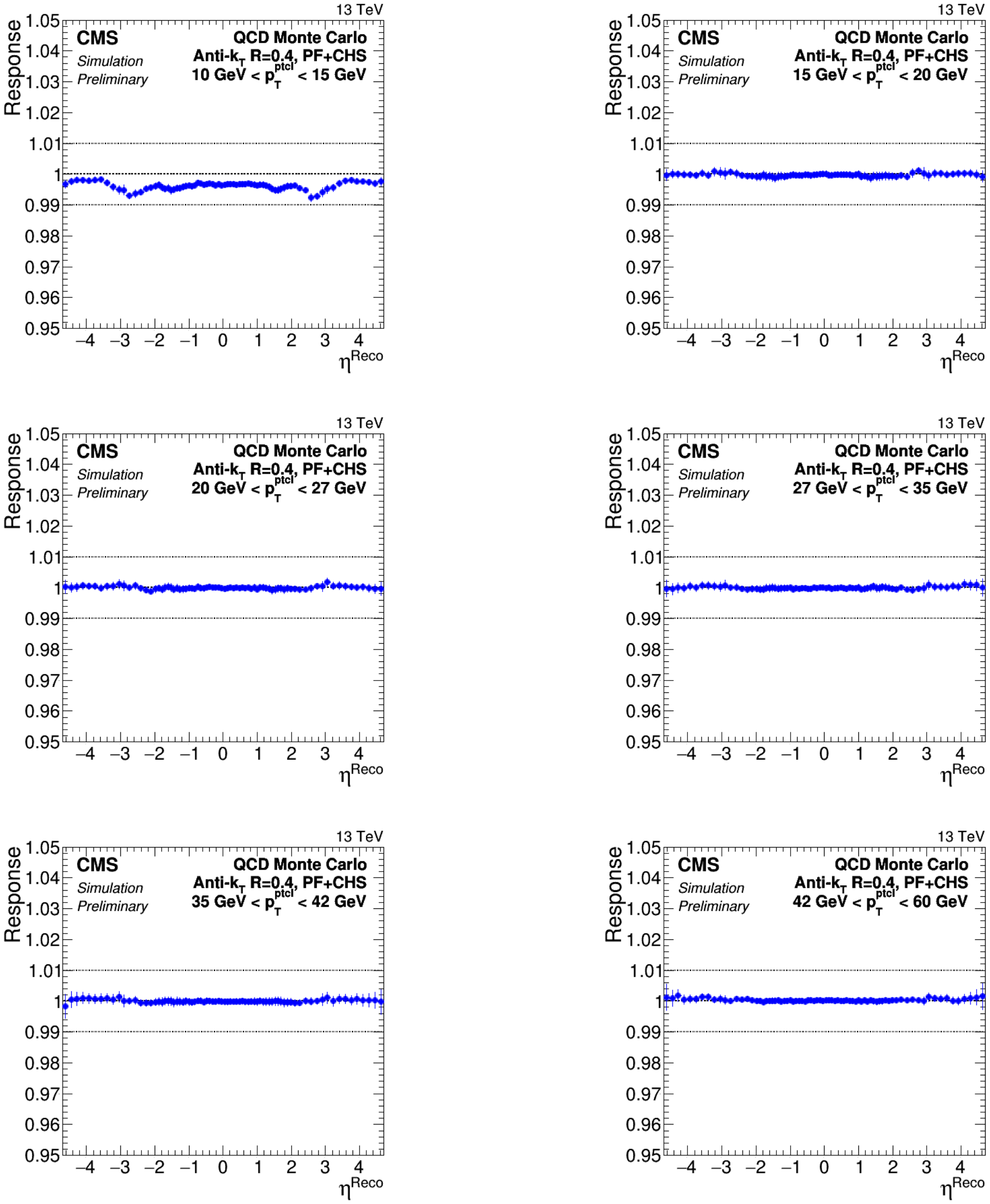


Figure C.1: The average of the response distributions in function of the reconstructed pseudorapidity η^{Reco} for the first 6 bins in p_T^{gen} for the closure test of the L2L3 MC-truth corrections.

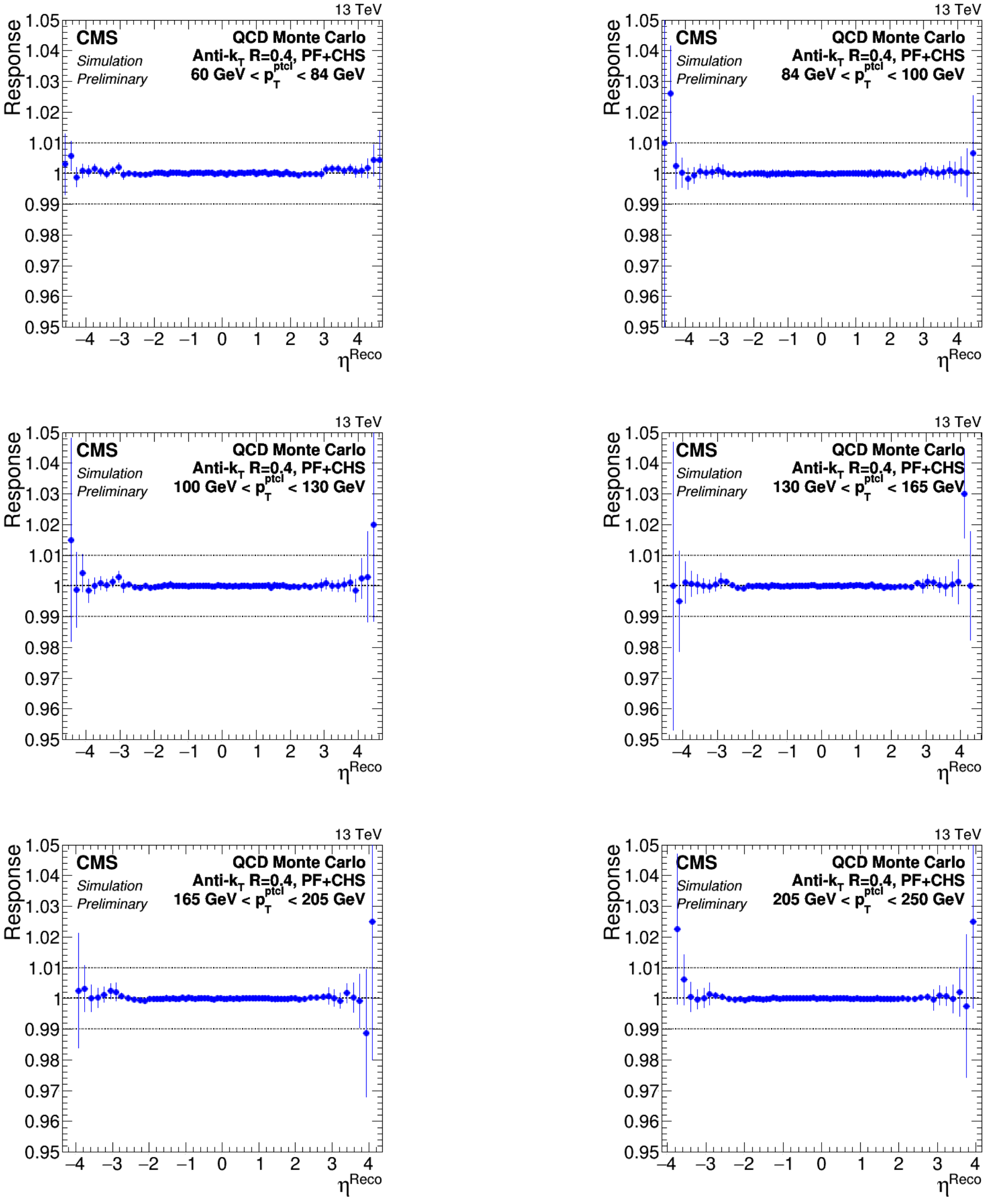
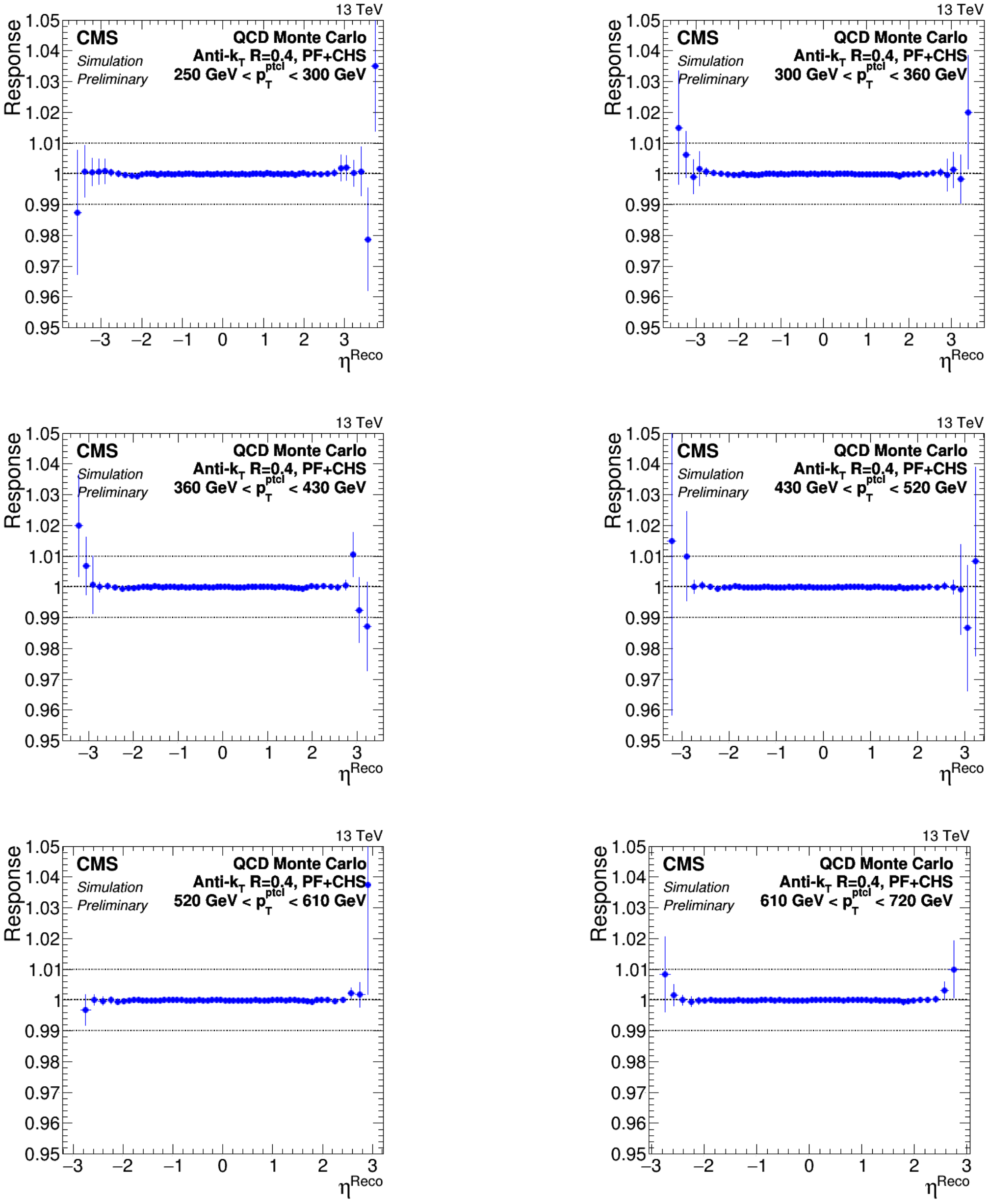


Figure C.2: The average of the response distributions in function of the reconstructed pseudorapidity η^{Reco} for the 7th up until the 12th bin in p_T^{gen} for the closure test of the L2L3 MC-truth corrections.



p

Figure C.3: The average of the response distributions in function of the reconstructed pseudorapidity η^{Reco} for the 13th up until the 18th bin in p_T^{gen} for the closure test of the L2L3 MC-truth corrections.

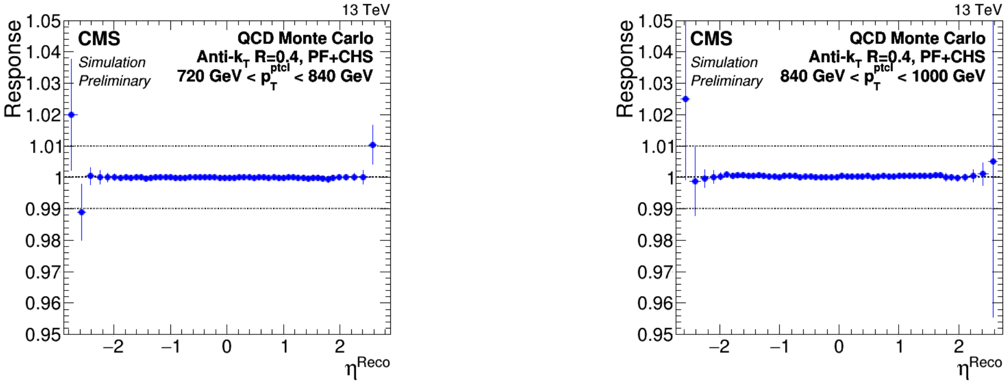


Figure C.4: The average of the response distributions in function of the reconstructed pseudorapidity η^{Reco} for the 19th and 20th bin in p_T^{gen} for the closure test of the L2L3 MC-truth corrections.

C.2 Distributions of A and B

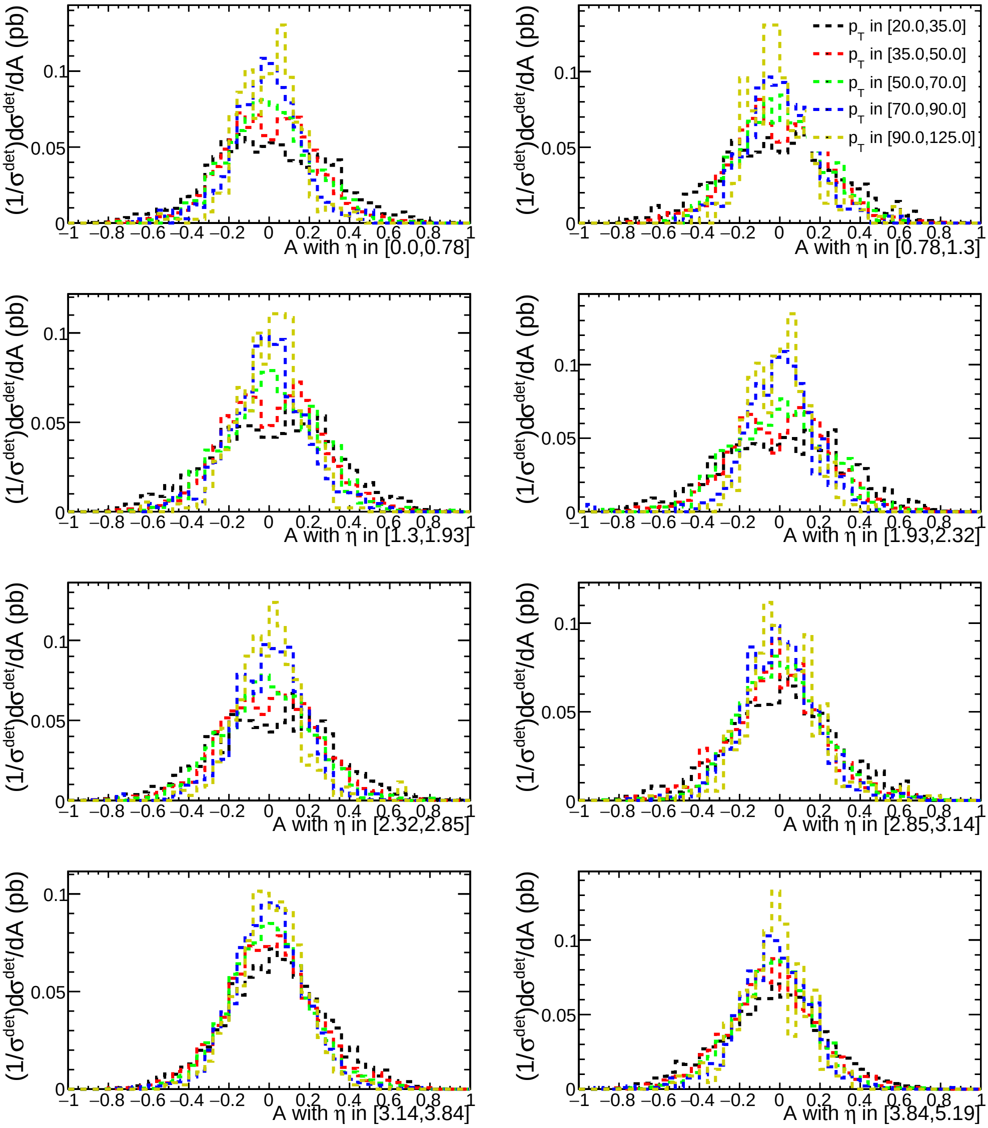


Figure C.5: Distributions of the variable A for each of the different p_{T} and η_{probe} bins for the `FSQJets` data sample.

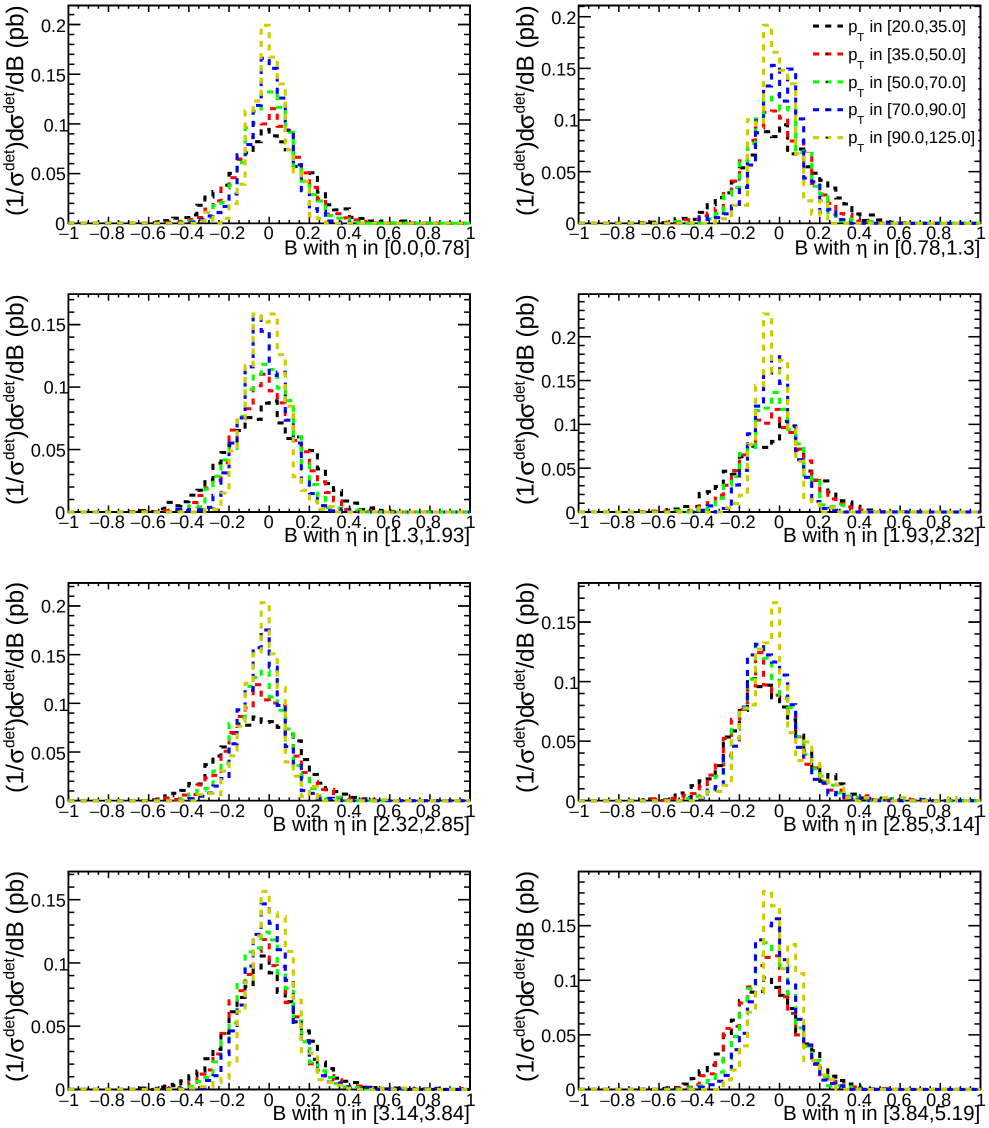


Figure C.6: Distributions of the variable B for each of the different p_T and η_probe bins for the `FSQJets` data sample.

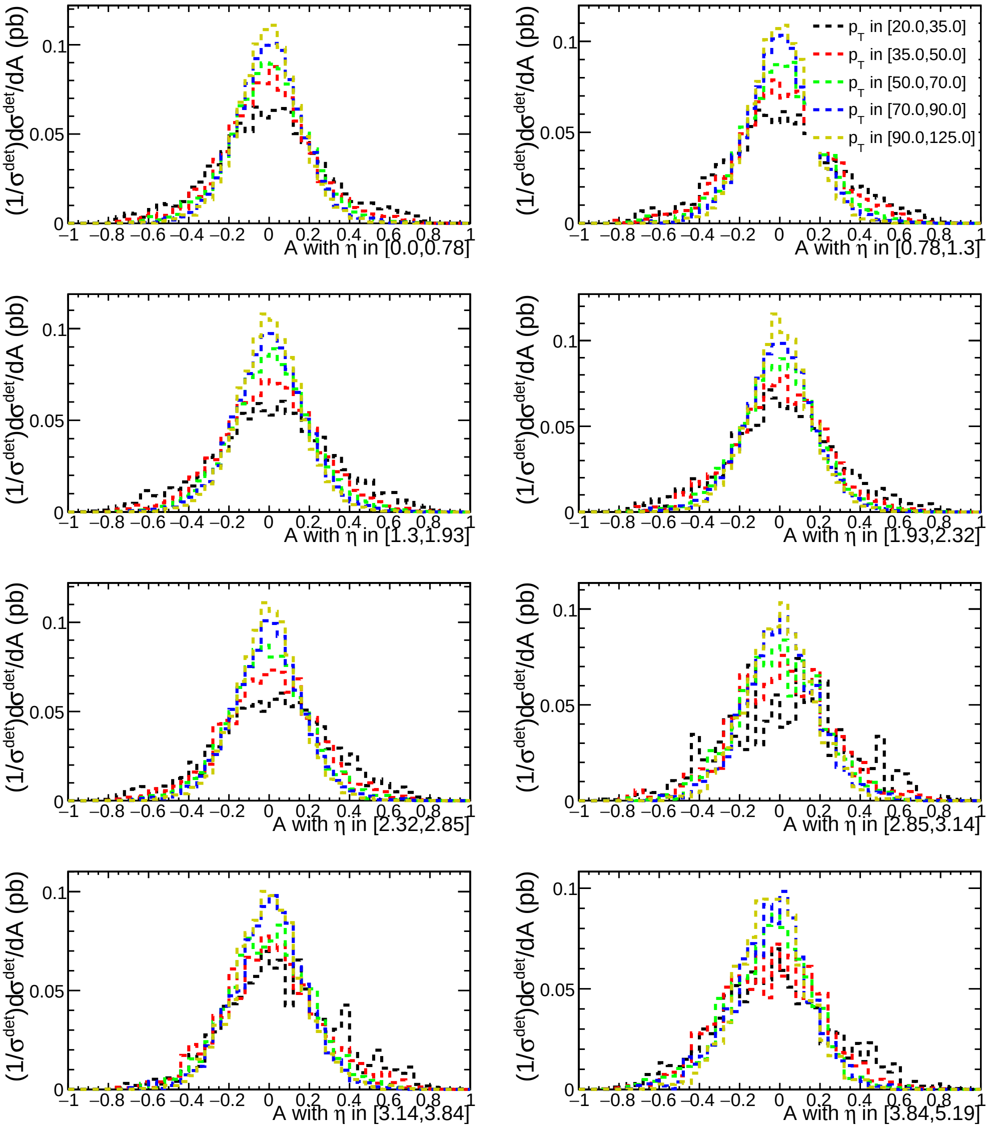


Figure C.7: Distributions of the variable A for each of the different p_{T} and η_{probe} bins for the `Pythia` 8 data sample.

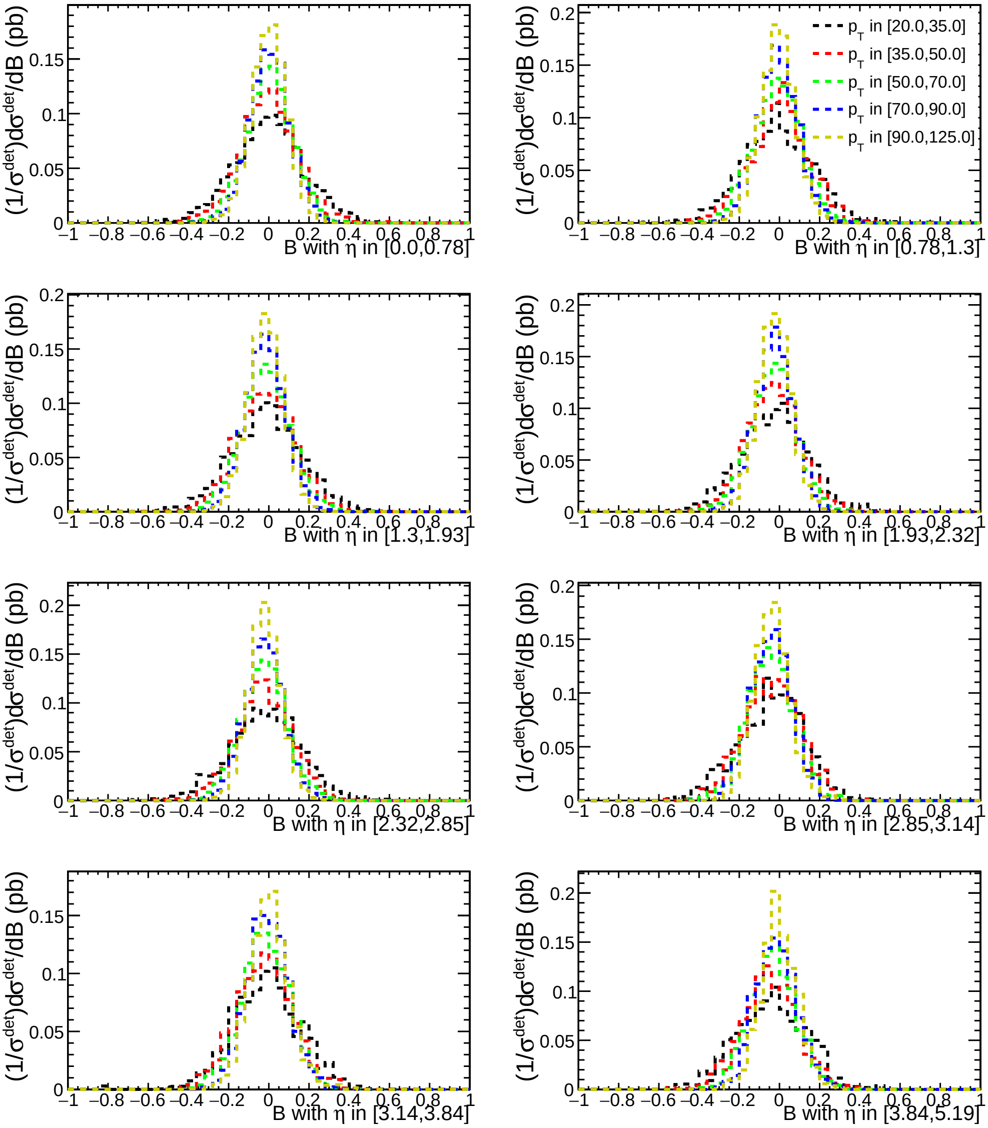


Figure C.8: Distributions of the variable B for each of the different p_{T} and η_{probe} bins for the `Pythia` 8 data sample.

C.3 Double Ratios of MC over Data

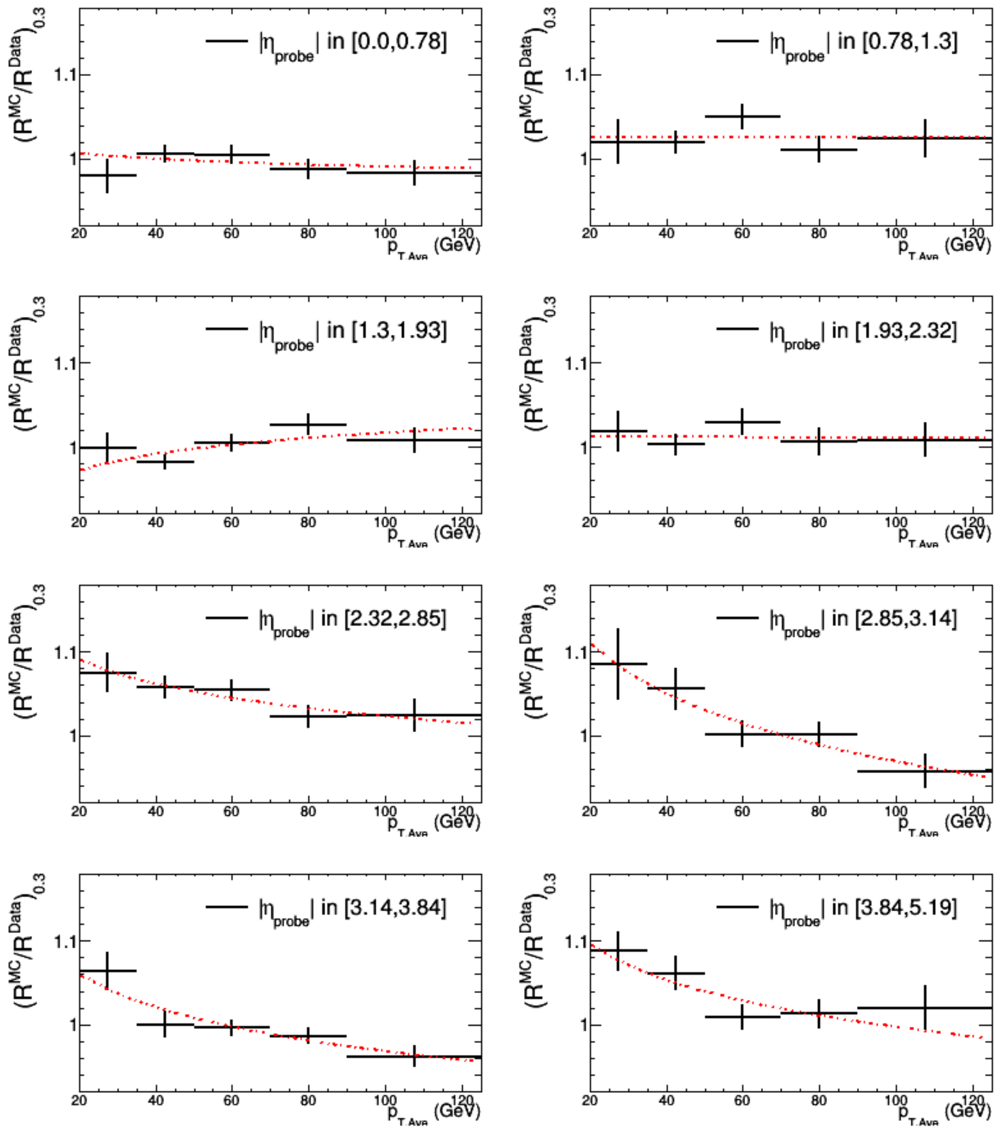


Figure C.9: The ratio $R_{\mathrm{rel}}^{\mathrm{MC}}/R_{\mathrm{rel}}^{\mathrm{DATA}}$ for the p_{T}-balance method in function of the $p_{\mathrm{T,Ave}}$ per bin in η_{probe}, where a maximum value of $\alpha_{\max} = 0.3$ has been used. A linear logarithmic function has been fitted in order to extract the p_{T} dependence is expected.

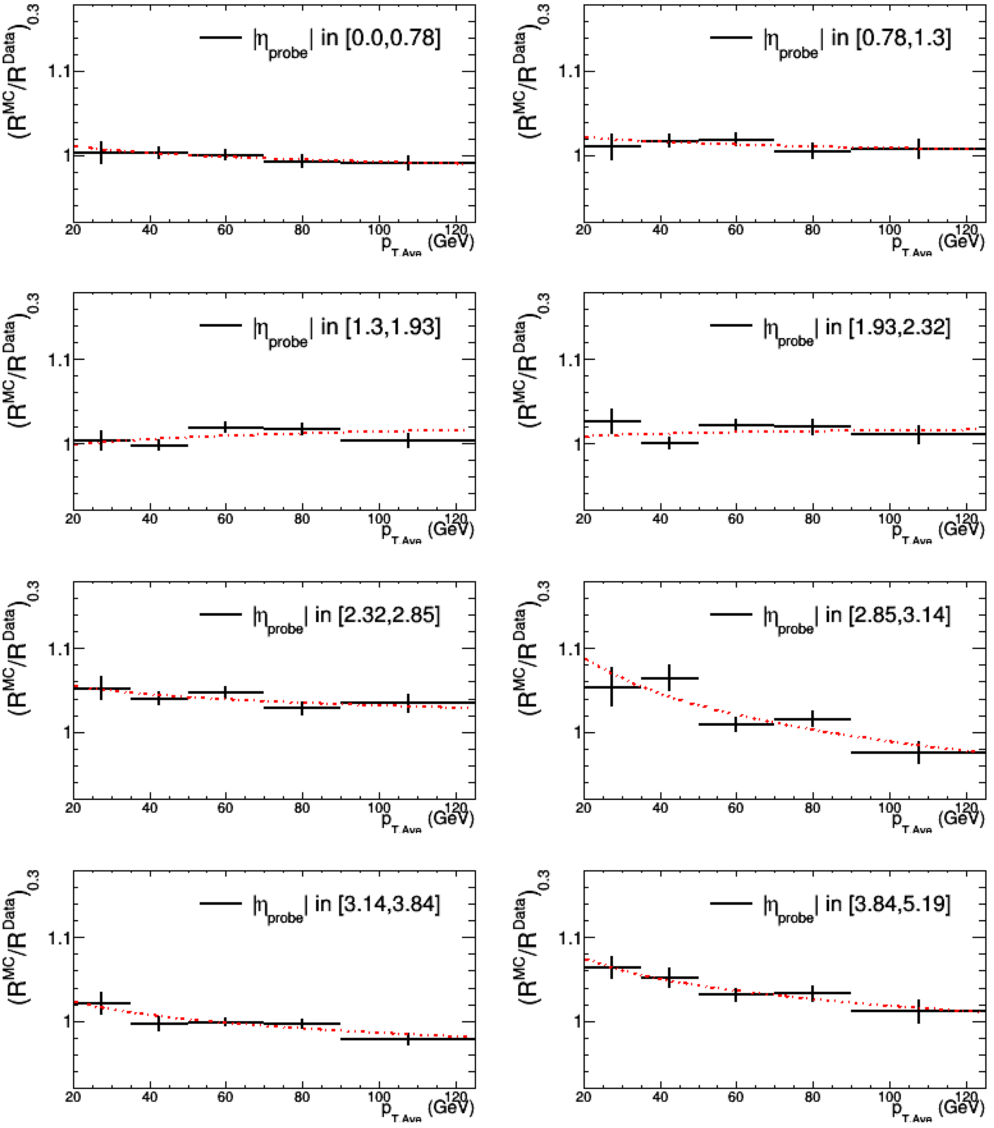


Figure C.10: The ratio $R^{\rm MC}_{\rm rel}/R^{\rm DATA}_{\rm rel}$ for the MPF method in function of the $p_{\rm T,Ave}$ per bin in $\eta_{\rm probe}$, where a maximum value of $\alpha_{\rm max} = 0.3$ has also been used. A linear logarithmic function has been fitted in order to extract the $p_{\rm T}$ dependence is expected.

C.4 The k_{FSR} Extrapolation

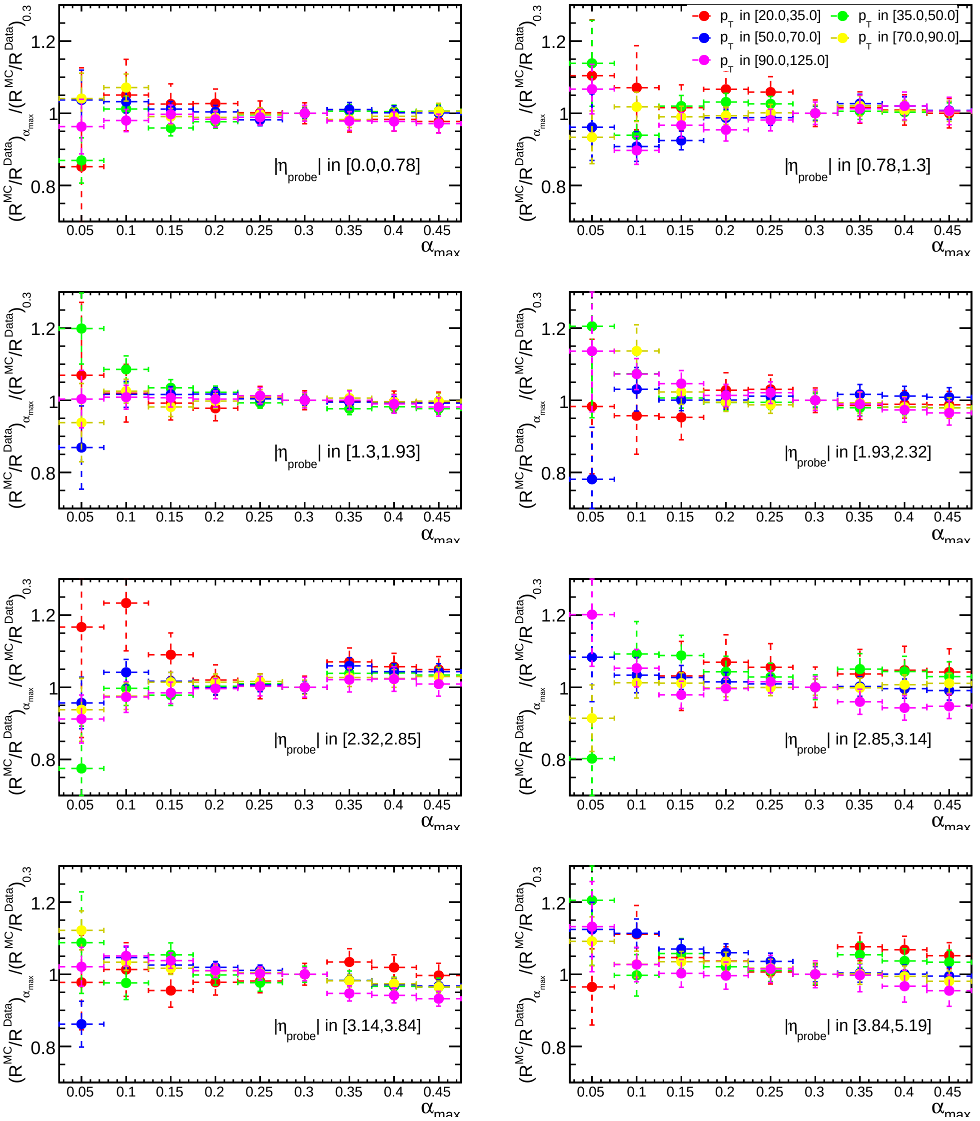


Figure C.11: Each plot represents a different rapidity bin, while the factor $k_{\mathrm{FSR}}(\alpha_{\mathrm{max}}, |\eta_{\mathrm{probe}}|)$ for the p_{T}-balance method is plotted for all p_{T} bins and in function of α_{max}

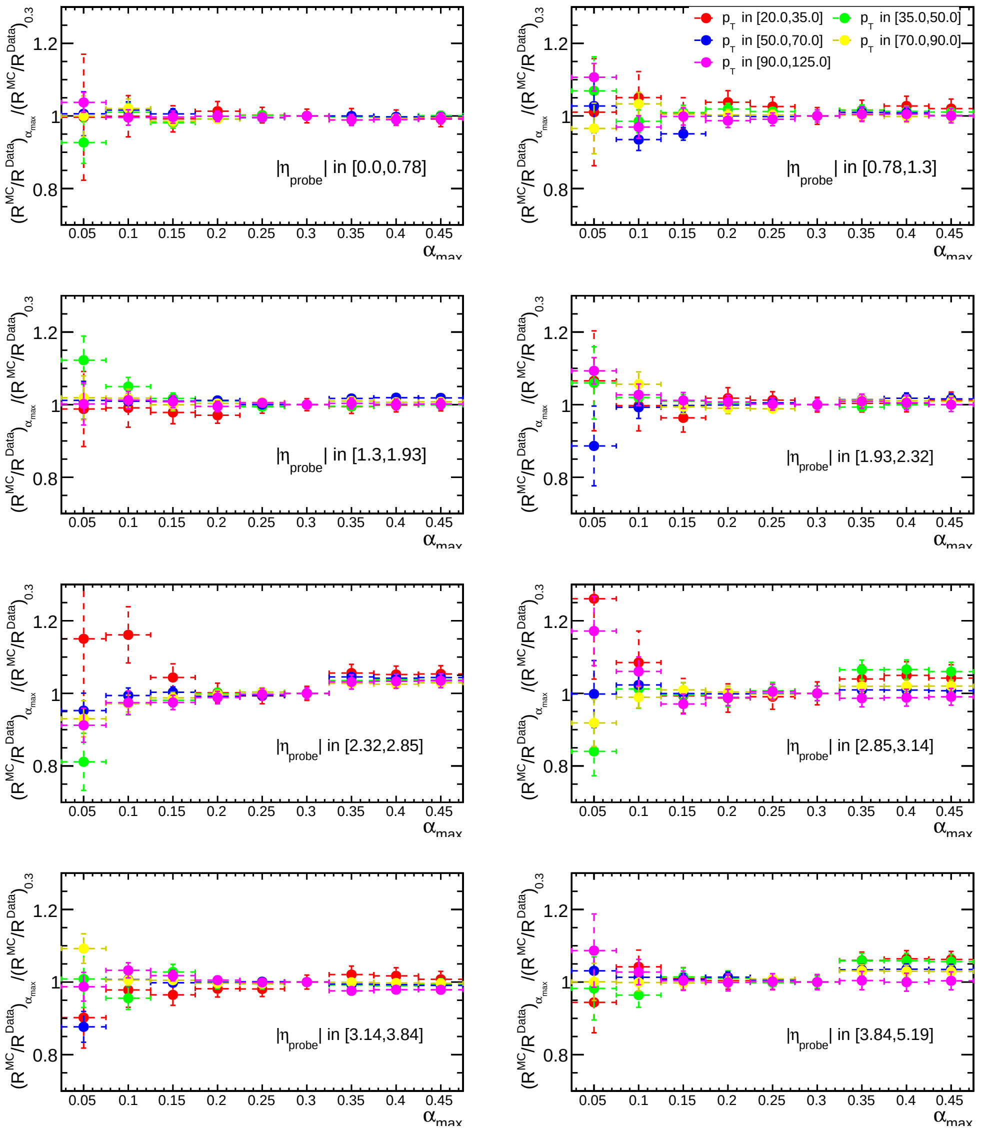


Figure C.12: Each plot represents a different rapidity bin, while the factor $k_{\mathrm{FSR}}(\alpha_{\max}, |\eta_{\mathrm{probe}}|)$ for the MPF method is plotted for all p_{T} bins and in function of $\alpha_{\max}$

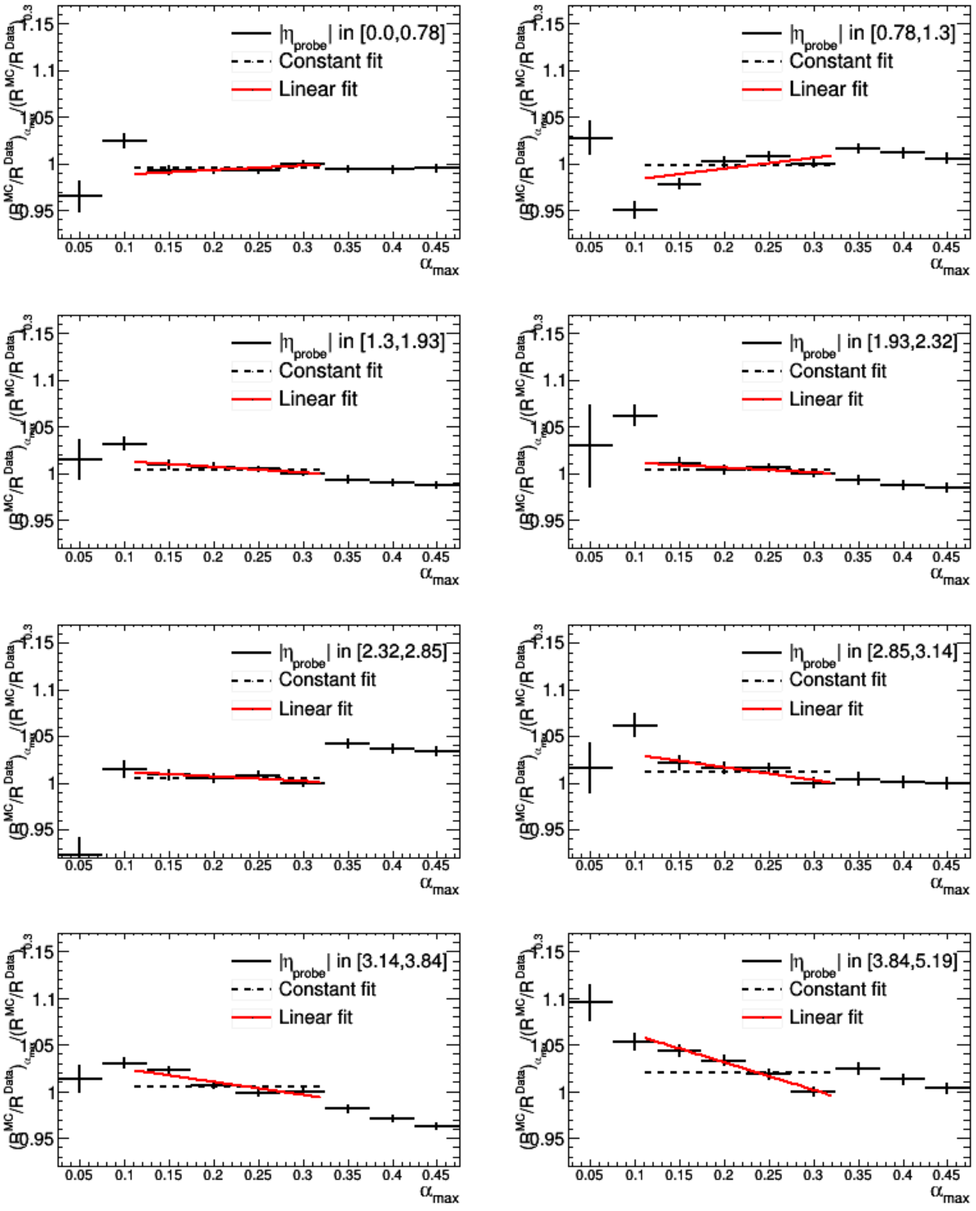


Figure C.13: Each plot represents a different rapidity bin, while the weighted average of the factor $k_{\mathrm{FSR}}(\alpha_{\max}, |\eta_{\mathrm{probe}}|)$ for the p_{T}-balance method is plotted. The weighted average over all p_{T}-bins for each value of $\alpha_{\max}$ is fitted with a constant and a linear fit for $\alpha_{\max} \in [0.11, 0.32]$.

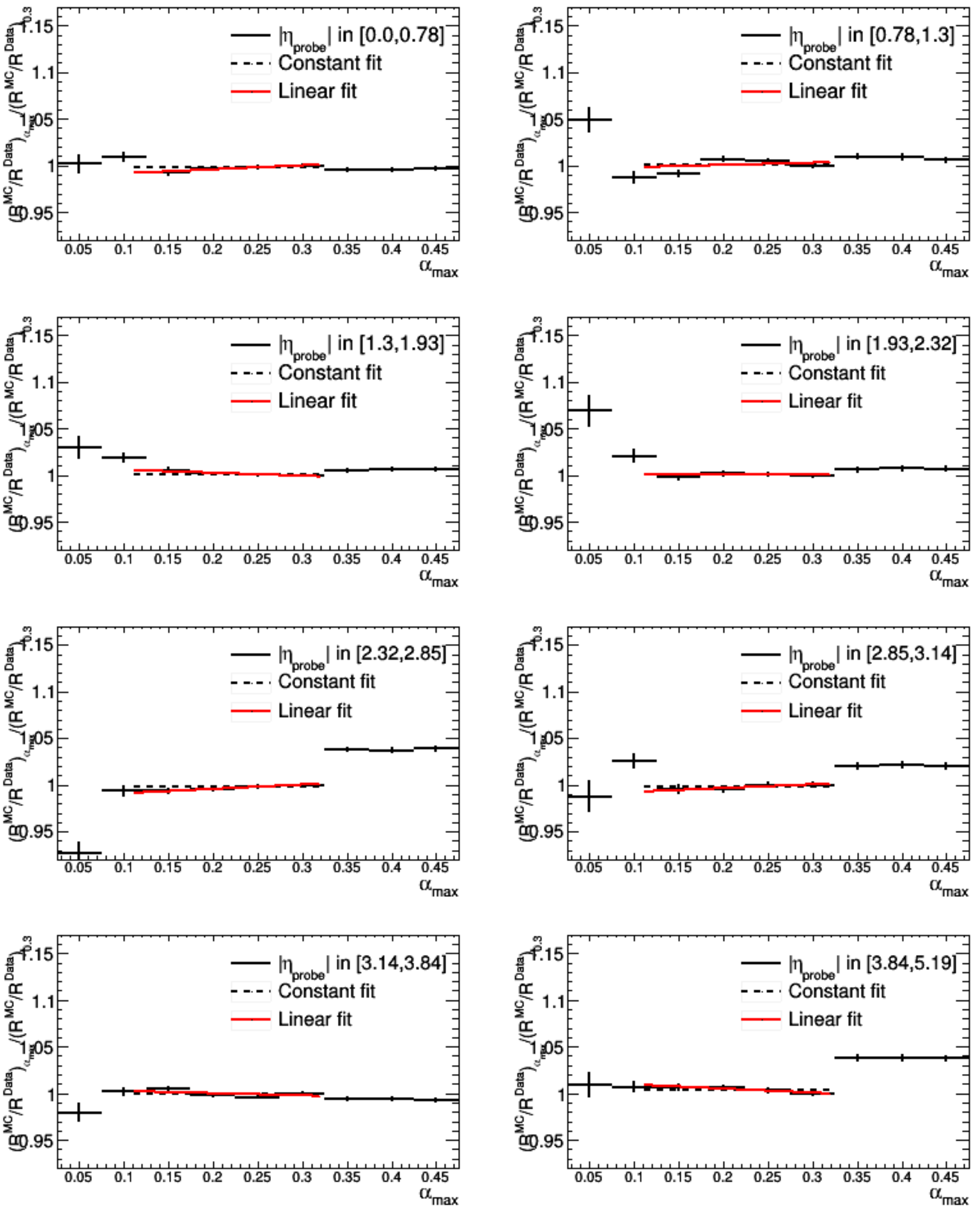


Figure C.14: Each plot represents a different rapidity bin, while the weighted average of the factor $k_{\mathrm{FSR}}(\alpha_{\max}, |\eta_{\mathrm{probe}}|)$ for the MPF method is plotted. The weighted average over all p_{T}-bins for each value of $\alpha_{\max}$ is fitted with a constant and a linear fit for $\alpha_{\max} \in [0.11, 0.32]$.

Bibliography

[1] R. Ellis and W. and Stirling. *An introduction to quantum field theory.* Westview, Boulder, CO, 1995. Includes exercises.

[2] Francis Halzen and Alan Martin. *Quarks & Leptons: An introductory course in modern particle physics.* John Wiley & Sons, New York, USA, 1984.

[3] Serguei Chatrchyan et al. Measurement of the Ratio of Inclusive Jet Cross Sections using the Anti-k_T Algorithm with Radius Parameters R=0.5 and 0.7 in pp Collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev.*, D90(7):072006, 2014.

[4] Vardan Khachatryan et al. Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV. *JHEP*, 03:156, 2017.

[5] Vardan Khachatryan et al. Measurement of the double-differential inclusive jet cross section in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *Eur. Phys. J.*, C76(8):451, 2016.

[6] F. Abe and others". Measurement of double parton scattering in $\overline{p}p$ collisions at $s = 1.8\text{TeV}$. *Phys. Rev. Lett.*, 79:584–589, Jul 1997.

[7] F. Abe and others". Double parton scattering in $\overline{p}p$ collisions at $\sqrt{s} = 1.8\text{TeV}$. *Phys. Rev. D*, 56:3811–3832, Oct 1997.

[8] V. M. Abazov and others". Double parton interactions in $\gamma + 3$ jet events in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev. D*, 81:052012, Mar 2010.

[9] Victor Mukhamedovich Abazov et al. Double Parton Interactions in $\gamma + 3$ Jet and $\gamma + b/cjet + 2$ Jet Events in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev.*, D89(7):072006, 2014.

[10] Victor Mukhamedovich Abazov et al. Study of double parton interactions in diphoton + dijet events in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev.*, D93(5):052008, 2016.

[11] F. Abe and others". Study of four-jet events and evidence for double parton interactions in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ tev. *Phys. Rev. D*, 47:4857–4871, Jun 1993.

[12] Georges Aad et al. Measurement of hard double-parton interactions in $W(\to l\nu)+$ 2 jet events at $\sqrt{s}$=7 TeV with the ATLAS detector. *New J. Phys.*, 15:033038, 2013.

[13] Serguei Chatrchyan et al. Study of Double Parton Scattering Using W + 2-Jet Events in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 03:032, 2014.

[14] Albert M Sirunyan et al. Constraints on the double-parton scattering cross section from same-sign W boson pair production in proton-proton collisions at $\sqrt{s} = 8$ TeV. *JHEP*, 02:032, 2018.

[15] Albert M Sirunyan et al. Evidence for WW production from double-parton interactions in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J.*, C80(1):41, 2020.

[16] R Aaij et al. Observation of double charm production involving open charm in pp collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 06:141, 2012. [Addendum: JHEP03,108(2014)].

[17] Roel Aaij et al. Observation of associated production of a Z boson with a D meson in the forward region. *JHEP*, 04:091, 2014.

[18] V. M. Abazov and others". Observation and studies of double j/ψ production at the tevatron. *Phys. Rev. D*, 90:111101, Dec 2014.

[19] Roel Aaij et al. Production of associated Y and open charm hadrons in pp collisions at $\sqrt{s} = 7$ and 8 TeV via double parton scattering. *JHEP*, 07:052, 2016.

[20] Roel Aaij et al. Measurement of the J/ψ pair production cross-section in pp collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 06:047, 2017.

[21] J. Alitti et al. A study of multi-jet events at the cern pp collider and a search for double parton scattering. *Physics Letters B*, 268(1):145 – 154, 1991.

[22] F. Abe et al. Study of four jet events and evidence for double parton interactions in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. *Phys. Rev. D*, 47:4857–4871, 1993.

[23] Georges Aad et al. Measurement of four-jet differential cross sections in $\sqrt{s} = 8$ TeV proton-proton collisions using the ATLAS detector. *JHEP*, 12:105, 2015.

[24] Iftach Sadeh. *Double parton scattering in four-jet events in pp collisions at 7 TeV with the ATLAS experiment at the LHC.* PhD thesis, Tel Aviv U., 2013-01-01.

[25] Morad Aaboud et al. Study of hard double-parton scattering in four-jet events in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS experiment. *JHEP*, 11:110, 2016.

[26] Serguei Chatrchyan et al. Measurement of four-jet production in proton-proton collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev.*, D89(9):092010, 2014.

[27] Vardan Khachatryan et al. Studies of inclusive four-jet production with two b-tagged jets in proton-proton collisions at 7 TeV. *Phys. Rev.*, D94(11):112005, 2016.

[28] J. J. Thomson M.A. F.R.S. Xl. cathode rays. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 44(269):293–316, 1897.

[29] W. Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift fur Physik*, 43(3-4):172–198, March 1927.

[30] J. Chadwick. Possible Existence of a Neutron. *Nature*, 129:312, 1932.

[31] E. Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Phys. Rev.*, 28:1049–1070, Dec 1926.

[32] C. N. Yang and R. L. Mills. Conservation of isotopic spin and isotopic gauge invariance. *Phys. Rev.*, 96:191–195, Oct 1954.

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

[34] Serguei Chatrchyan et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012.

[35] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[36] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[37] Wikipedia contributors. Standard model — Wikipedia, the free encyclopedia, 2020. [Online; accessed 15-May-2020].

[38] A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 322(8):549–560, January 1905.

[39] O Klein. Quantentheorie und fünfdimensionale relativitätstheorie. *Z. Phys.*, 37:895–906, 1926.

[40] W. Gordon. Der Comptoneffekt nach der Schrödingerschen Theorie. *Zeitschrift fur Physik*, 40(1-2):117–133, January 1926.

[41] P. A. M. Dirac. The Quantum Theory of the Electron. *Proceedings of the Royal Society of London Series A*, 117(778):610–624, February 1928.

[42] Sheldon L. Glashow. The renormalizability of vector meson interactions. *Nucl. Phys.*, 10:107–117, 1959.

[43] Abdus Salam and John Clive Ward. Weak and electromagnetic interactions. *Nuovo Cim.*, 11:568–577, 1959.

[44] Steven Weinberg. A Model of Leptons. , 19(21):1264–1266, November 1967.

[45] Albert M. Sirunyan et al. Measurement of the weak mixing angle using the forward-backward asymmetry of Drell-Yan events in pp collisions at 8 TeV. *Eur. Phys. J. C*, 78(9):701, 2018.

[46] Wikipedia contributors. Spontaneous symmetry breaking — Wikipedia, the free encyclopedia, 2020. [Online; accessed 19-May-2020].

[47] Hideki Yukawa. On the Interaction of Elementary Particles I. *Proc. Phys. Math. Soc. Jap.*, 17:48–57, 1935.

[48] Albert Einstein. Cosmological Considerations in the General Theory of Relativity. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*, 1917:142–152, 1917.

[49] Stephen P. Martin. *A Supersymmetry primer*, volume 21, pages 1–153. 2010.

[50] Laurent Canetti, Marco Drewes, and Mikhail Shaposhnikov. Matter and Antimatter in the Universe. *New J. Phys.*, 14:095012, 2012.

[51] V. Barger, D. Marfatia, and K. Whisnant. *The Physics of Neutrinos.* Princeton University Press, 2012.

[52] Virginia Trimble. Existence and nature of dark matter in the universe. , 25:425–472, January 1987.

[53] P.J.E. Peebles and Bharat Ratra. The Cosmological Constant and Dark Energy. *Rev. Mod. Phys.*, 75:559–606, 2003.

[54] Gerard 't Hooft. Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking. *NATO Sci. Ser. B*, 59:135–157, 1980.

[55] G.M. Prosperi, M. Raciti, and C. Simolo. On the running coupling constant in QCD. *Prog. Part. Nucl. Phys.*, 58:387–438, 2007.

[56] David J. Gross and Frank Wilczek. Ultraviolet behavior of non-abelian gauge theories. *Phys. Rev. Lett.*, 30:1343–1346, Jun 1973.

[57] H. David Politzer. Reliable perturbative results for strong interactions? *Phys. Rev. Lett.*, 30:1346–1349, Jun 1973.

[58] Elliott D. Bloom et al. High-Energy Inelastic e p Scattering at 6-Degrees and 10-Degrees. *Phys. Rev. Lett.*, 23:930–934, 1969.

[59] T. Ahmed et al. A Measurement of the proton structure function f2 (x, Q**2). *Nucl. Phys. B*, 439:471–502, 1995.

[60] C. Gwenlan. Combined hera deep inelastic scattering data and nlo qcd fits. *Nuclear Physics B - Proceedings Supplements*, 191:5–15, Jun 2009.

[61] Edmond Iancu. QCD in heavy ion collisions. (arXiv:1205.0579):197–266. 73 p, May 2012. Based on lectures presented at the 2011 European School of High-Energy Physics, 7-20 September 2011, Cheile Gradistei, Romania. 73 pages, many figures.

[62] G. Altarelli and G. Parisi. Asymptotic freedom in parton language. *Nuclear Physics B*, 126(2):298 – 318, 1977.

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

[64] S. Moch, B. Ruijl, T. Ueda, J.A.M. Vermaseren, and A. Vogt. Four-Loop Non-Singlet Splitting Functions in the Planar Limit and Beyond. *JHEP*, 10:041, 2017.

[65] V V Sudakov. Vertex parts at very high energies in quantum electrodynamics. *Zh. Eksp. Teor. Fiz.*, 30:87–95, 1956.

[66] E. A. Kuraev, L. N. Lipatov, and Victor S. Fadin. Multi - Reggeon Processes in the Yang-Mills Theory. *Sov. Phys. JETP*, 44:443–450, 1976.

[67] E.A. Kuraev, L.N. Lipatov, and Victor S. Fadin. The Pomeranchuk Singularity in Nonabelian Gauge Theories. *Sov. Phys. JETP*, 45:199–204, 1977.

[68] Francois Gelis, Edmond Iancu, Jamal Jalilian-Marian, and Raju Venugopalan. The Color Glass Condensate. *Ann. Rev. Nucl. Part. Sci.*, 60:463–489, 2010.

[69] S. Catani, F. Fiorani, and G. Marchesini. Qcd coherence in initial state radiation. *Physics Letters B*, 234(3):339 – 345, 1990.

[70] S. Catani, F. Fiorani, and G. Marchesini. Small-x behaviour of initial state radiation in perturbative qcd. *Nuclear Physics B*, 336(1):18 – 85, 1990.

[71] G. Gustafson. Dual description of a confined colour field. *Physics Letters B*, 175(4):453 – 456, 1986.

[72] Gösta Gustafson and Ulf Pettersson. Dipole formulation of qcd cascades. *Nuclear Physics B*, 306(4):746 – 758, 1988.

[73] Bo Andersson, Gösta Gustafson, and Leif Lönnblad. Gluon splitting in the colour dipole cascades. *Nuclear Physics B*, 339(2):393 – 405, 1990.

[74] John Collins. *Foundations of Perturbative QCD.* Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 2011.

[75] S. Catani and F. Hautmann. High-energy factorization and small x deep inelastic scattering beyond leading order. *Nucl. Phys.*, B427:475–524, 1994.

[76] M. Deak, F. Hautmann, H. Jung, and K. Kutak. Forward Jet Production at the Large Hadron Collider. *JHEP*, 09:121, 2009.

[77] Sebastian Sapeta. QCD and Jets at Hadron Colliders. *Prog. Part. Nucl. Phys.*, 89:1–55, 2016.

[78] Andy Buckley et al. General-purpose event generators for LHC physics. *Phys. Rept.*, 504:145–233, 2011.

[79] Stefan Hoeche, Frank Krauss, Nils Lavesson, Leif Lonnblad, Michelangelo Mangano, Andreas Schalicke, and Steffen Schumann. Matching parton showers and matrix elements. In *HERA and the LHC: A Workshop on the implications of HERA for LHC physics: Proceedings Part A*, pages 288–289, 2005.

[80] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[81] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

[82] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[83] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. 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.

[84] Johan Alwall et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J.*, C53:473–500, 2008.

[85] Frank Siegert. Sherpa wiki page. 2016.

[86] Albert M Sirunyan et al. Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV. *JHEP*, 07:161, 2018.

[87] G. Antchev et al. First measurement of elastic, inelastic and total cross-section at $\sqrt{s} = 13$ TeV by TOTEM and overview of cross-section data at LHC energies. *Eur. Phys. J. C*, 79(2):103, 2019.

[88] M. Aaboud et al. Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC. *Phys. Rev. Lett.*, 117(18):182002, 2016.

[89] Underlying Event Measurements with Leading Particles and Jets in pp collisions at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-FSQ-15-007, CERN, Geneva, 2015.

[90] Measurement of the underlying event using the Drell-Yan process in proton-proton collisions at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-FSQ-16-008, CERN, Geneva, 2016.

[91] Underlying event measurement with ttbar+X events with p-p collision data at sqrt(s)=13 TeV. Technical Report CMS-PAS-TOP-15-017, CERN, Geneva, 2015.

[92] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

[93] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.

[94] M. Bahr et al. Herwig++ Physics and Manual. *Eur. Phys. J.*, C58:639–707, 2008.

[95] Johannes Bellm et al. Herwig 7.0 / Herwig++ 3.0 Release Note. *Eur. Phys. J. C*, 76(CERN-PH-TH-2015-289. CERN-PH-TH-2015-289. MAN-HEP-2015-15. IFJPAN-IV-2015-13. HERWIG-2015-01. KA-TP-18-2015. DCPT-15-142. MCNET-15-28. IPPP-15-71):196. 8 p, Dec 2015. 8 pages. Herwig is available from https://herwig.hepforge.org.

[96] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand. Parton fragmentation and string dynamics. *Physics Reports*, 97(2):31 – 145, 1983.

[97] G. Marchesini and B. R. Webber. Simulation of QCD Jets Including Soft Gluon Interference. *Nucl. Phys.*, B238:1–29, 1984.

[98] B. R. Webber. A QCD Model for Jet Fragmentation Including Soft Gluon Interference. *Nucl. Phys.*, B238:492–528, 1984.

[99] G. Marchesini and B. R. Webber. Monte Carlo Simulation of General Hard Processes with Coherent QCD Radiation. *Nucl. Phys.*, B310:461–526, 1988.

[100] Bryan R Webber. Hadronization. (hep-ph/9411384. CAVENDISH-HEP-94-17), 1994.

[101] D. Amati and G. Veneziano. Preconfinement as a property of perturbative qcd. *Physics Letters B*, 83(1):87 – 92, 1979.

[102] B.R. Webber. Fragmentation and Hadronization. *Int. J. Mod. Phys. AS*, 15S1(hep-ph/9912292. CAVENDISH-HEP-99-16. CERN-TH-99-387):577–606. 31 p, Jan 2000.

[103] Gerald C. Blazey et al. Run II jet physics. In *Physics at Run II: QCD and Weak Boson Physics Workshop: Final General Meeting*, pages 47–77, 5 2000.

[104] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-ktjet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, apr 2008.

[105] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D*, 48:3160–3166, 1993.

[106] Yuri L. Dokshitzer, G.D. Leder, S. Moretti, and B.R. Webber. Better jet clustering algorithms. *JHEP*, 08:001, 1997.

[107] Markus Diehl. Correlation effects in multiple hard scattering. *PoS*, DIS2013:074, 2013.

[108] Michael H. Seymour and Andrzej Siodmok. Extracting $\sigma_{\text{effective}}$ from the LHCb double-charm measurement. 8 2013.

[109] Markus Diehl, Daniel Ostermeier, and Andreas Schafer. Elements of a theory for multiparton interactions in QCD. *JHEP*, 03:089, 2012. [Erratum: JHEP03,001(2016)].

[110] Markus Diehl, Tomas Kasemets, and Shane Keane. Correlations in double parton distributions: effects of evolution. *JHEP*, 05:118, 2014.

[111] M. Rinaldi, S. Scopetta, and V. Vento. Double parton correlations in constituent quark models. *Phys. Rev.*, D87:114021, 2013.

[112] Maarten G. A. Buffing, Markus Diehl, and Tomas Kasemets. Transverse momentum in double parton scattering: factorisation, evolution and matching. *JHEP*, 01:044, 2018.

[113] R. K. Ellis, W. J. Stirling, and B. R. Webber. *QCD and Collider Physics*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 1996.

[114] Svend Domdey, Hans-Jurgen Pirner, and Urs Achim Wiedemann. Testing the Scale Dependence of the Scale Factor sigma(eff) in Double Dijet Production at the LHC. *Eur. Phys. J.*, C65:153–162, 2010.

[115] Edmond L. Berger, C. B. Jackson, and Gabe Shaughnessy. Characteristics and Estimates of Double Parton Scattering at the Large Hadron Collider. *Phys. Rev.*, D81:014014, 2010.

[116] Rafał Maciuła and Antoni Szczurek. Searching for and exploring double-parton scattering effects in four-jet production at the LHC. *Phys. Lett.*, B749:57–62, 2015.

[117] Krzysztof Kutak, Rafal Maciula, Mirko Serino, Antoni Szczurek, and Andreas van Hameren. Four-jet production in single- and double-parton scattering within high-energy factorization. *JHEP*, 04:175, 2016.

[118] Krzysztof Kutak. Production of four-jets in single- and double-parton scattering within high-energy factorization. *PoS*, ICHEP2016:610, 2016.

[119] Krzysztof Kutak, Rafał Maciuła, Mirko Serino, Antoni Szczurek, and Andreas van Hameren. Search for optimal conditions for exploring double-parton scattering in four-jet production: k_t-factorization approach. *Phys. Rev.*, D94(1):014019, 2016.

[120] Peter Skands, Stefano Carrazza, and Juan Rojo. Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur. Phys. J.*, C74(8):3024, 2014.

[121] Vardan Khachatryan et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J.*, C76(3):155, 2016.

[122] Richard D. Ball, Valerio Bertone, Francesco Cerutti, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Jose I. Latorre, Juan Rojo, and Maria Ubiali. Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO. *Nucl. Phys.*, B855:153–221, 2012.

[123] Stefano Carrazza, Stefano Forte, and Juan Rojo. Parton Distributions and Event Generators. In *Proceedings, 43rd International Symposium on Multiparticle Dynamics (ISMD 13)*, pages 89–96, 2013.

[124] Albert M Sirunyan et al. Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements. 2019.

[125] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J.*, C77(10):663, 2017.

[126] Richard Corke and Torbjorn Sjostrand. Interleaved Parton Showers and Tuning Prospects. *JHEP*, 03:032, 2011.

[127] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, Pavel M. Nadolsky, and W. K. Tung. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 07:012, 2002.

[128] Nadine Fischer, Stefan Prestel, Mathias Ritzmann, and Peter Skands. Vincia for Hadron Colliders. *Eur. Phys. J.*, C76(11):589, 2016.

[129] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne. Parton distributions in the LHC era: MMHT 2014 PDFs. *Eur. Phys. J.*, C75(5):204, 2015.

[130] H. Jung et al. The CCFM Monte Carlo generator CASCADE version 2.2.03. *Eur. Phys. J.*, C70:1237–1249, 2010.

[131] Simone Alioli, Keith Hamilton, Paolo Nason, Carlo Oleari, and Emanuele Re. Jet pair production in POWHEG. *JHEP*, 04:081, 2011.

[132] Adam Kardos, Paolo Nason, and Carlo Oleari. Three-jet production in POWHEG. *JHEP*, 04:043, 2014.

[133] A. van Hameren. KaTie : For parton-level event generation with k_T-dependent initial states. *Comput. Phys. Commun.*, 224:371–380, 2018.

[134] Marcin Bury, Andreas van Hameren, Hannes Jung, Krzysztof Kutak, Sebastian Sapeta, and Mirko Serino. Calculations with off-shell matrix elements, TMD parton densities and TMD parton showers. *Eur. Phys. J.*, C78(2):137, 2018.

[135] A. Bermudez Martinez, P. Connor, H. Jung, A. Lelek, R. Žlebčík, F. Hautmann, and V. Radescu. Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method. *Phys. Rev.*, D99(7):074008, 2019.

[136] T. Sjostrand and Peter Z. Skands. Transverse-momentum-ordered showers and interleaved multiple interactions. *Eur. Phys. J. C*, 39:129–154, 2005.

[137] Richard Corke and Torbjorn Sjostrand. Multiparton Interactions and Rescattering. *JHEP*, 01:035, 2010.

[138] Stefan Gieseke, Christian Rohr, and Andrzej Siodmok. Multiple Partonic Interactions in Herwig++. *Acta Phys. Polon. Supp.*, 6:613–620, 2013.

[139] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008.

[140] *ATLAS: technical proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN.* LHC Tech. Proposal. CERN, Geneva, 1994.

[141] *Technical proposal.* LHC Tech. Proposal. CERN, Geneva, 1994. Cover title : CMS, the Compact Muon Solenoid : technical proposal.

[142] S Chatrchyan et al. The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment. *JINST*, 3:S08004. 361 p, 2008. Also published by CERN Geneva in 2010.

[143] *LHCb : Technical Proposal.* Tech. Proposal. CERN, Geneva, 1998.

[144] W et al. Kienzle. TOTEM, Total Cross Section, Elastic Scattering and Diffraction Dissociation at the LHC: Technical Proposal. Technical Report CERN-LHCC-99-007. LHCC-P-5, CERN, Geneva, Mar 1999.

[145] *ALICE: Technical proposal for a Large Ion collider Experiment at the CERN LHC.* LHC Tech. Proposal. CERN, Geneva, 1995.

[146] Jean-Luc Caron. Accelerator complex of CERN: an overview of all accelerators of CERN. Vue d'ensemble de tous les accélérateurs du CERN. Jul 2001.

[147] Tai Sakuma. Cutaway diagrams of CMS detector. May 2019.

[148] *The CMS tracker: addendum to the Technical Design Report.* Technical Design Report CMS. CERN, Geneva, 2000.

[149] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.

[150] *The CMS magnet project: Technical Design Report.* Technical Design Report CMS. CERN, Geneva, 1997.

[151] *The CMS electromagnetic calorimeter project: Technical Design Report.* Technical Design Report CMS. CERN, Geneva, 1997.

[152] Serguei et al. Chatrchyan. Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV. *JINST*, 8(CMS-EGM-11-001. CMS-EGM-11-001. CERN-PH-EP-2013-097):P09009. 51 p, Jun 2013. Comments: Submitted to JINST.

[153] *The CMS hadron calorimeter project: Technical Design Report.* Technical Design Report CMS. CERN, Geneva, 1997. The following files are from http://uscms.fnal.gov/pub/hcal_tdr and may not be the version as printed, please check the printed version to be sure.

[154] CMS Collaboration. Performance of CMS hadron calorimeter timing and synchronization using test beam, cosmic ray, and LHC beam data. *Journal of Instrumentation*, 5(03):T03013–T03013, mar 2010.

[155] CMS Collaboration. Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data. *Journal of Instrumentation*, 5(03):T03012–T03012, mar 2010.

[156] G et al. Baiatian. Design, Performance and Calibration of the CMS Forward Calorimeter Wedges. Technical Report CMS-NOTE-2006-044, CERN, Geneva, Feb 2006.

[157] J. G. Layter. *The CMS muon project: Technical Design Report.* Technical Design Report CMS. CERN, Geneva, 1997.

[158] Serguei Chatrchyan et al. The Performance of the CMS Muon Detector in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV at the LHC. *JINST*, 8:P11002, 2013.

[159] Serguei Chatrchyan et al. Performance of CMS Muon Reconstruction in pp Collision Events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012.

[160] A. Tapper and Darin Acosta. CMS Technical Design Report for the Level-1 Trigger Upgrade. 2013.

[161] Olivier Davignon. Design and performance of the upgrade of the CMS L1 trigger. *PoS*, ICHEP2018:198, 2019.

[162] Sergio Cittolin, Attila Rácz, and Paris Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project.* Technical Design Report CMS. CERN, Geneva, 2002.

[163] Laurent Thomas. CMS High Level Trigger performance at 13 TeV. *PoS*, ICHEP2018:226, 2019.

[164] John Stupak and Nitish Dhingra. Cms twiki page on dqm. 2013.

[165] Vardan Khachatryan et al. The CMS trigger system. *JINST*, 12(01):P01020, 2017.

[166] CMS Collaboration. CMS Luminosity Measurements for the 2016 Data Taking Period. 2017.

[167] S. Agostinelli et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003.

[168] Véronique Lefébure, Sudeshna Banerjee, and I. González. CMS Simulation Software Using Geant4. 1999.

[169] A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017.

[170] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet user manual. *Eur. Phys. J. C*, 72:1896, 2012.

[171] Albert M Sirunyan et al. Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector. *JINST*, 14(07):P07004, 2019.

[172] Vardan Khachatryan et al. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST*, 12(02):P02014, 2017.

[173] Anastasia Karavdina et al. Determination of the residual jet-energy corrections with dijet events at 13 TeV. 2016.

[174] The CMS collaboration. Determination of jet energy calibration and transverse momentum resolution in CMS. *Journal of Instrumentation*, 6(11):P11002–P11002, nov 2011.

[175] F. Abe et al. Dijet angular distribution in pp collisions at $\sqrt{\mathrm{s}}$ =1.8 TeV. *Phys. Rev. Lett.*, 69:2896–2900, Nov 1992.

[176] Rene Brun and Fons Rademakers. Root — an object oriented data analysis framework. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 389(1):81 – 86, 1997. New Computing Techniques in Physics Research V.

[177] Andreas Höcker and Vakhtang Kartvelishvili. Svd approach to data unfolding. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 372(3):469 – 481, 1996.

[178] A. N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.*, 4:1035–1038, 1963.

[179] Stefan Schmitt. TUnfold: an algorithm for correcting migration effects in high energy physics. *JINST*, 7:T10003, 2012.

[180] Stefan Schmitt. Data Unfolding Methods in High Energy Physics. *EPJ Web Conf.*, 137:11008, 2017.

[181] Stefan Schmitt. Tunfold package: user manual. 2018.

[182] CMS Luminosity Measurements for the 2016 Data Taking Period. Technical Report CMS-PAS-LUM-17-001, CERN, Geneva, 2017.

[183] J. Pekkanen. Jet energy resolution twiki pake. 2019.

[184] Inelastic pp cross section at 7 TeV. Technical Report CMS-PAS-FWD-11-001, CERN, Geneva, 2011.