IEKP-KA/2016-05

Karlsruher Institut für Technologie

Measurement of Triple-Differential Dijet Cross Sections with the CMS Detector at 8 TeV and PDF Constraints

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von

Dipl.-Phys. Georg Sieber
aus Berlin

Tag der mündlichen Prüfung: 20. Mai 2016

Referent: Prof. Dr. Günter Quast
Korreferent: Prof. Dr. Thomas Müller

Abstract

The first measurement of triple-differential dijet cross sections at the LHC is presented using 19.71 fb^{-1} of data collected with the CMS detector in proton-proton collisions at 8 TeV. The cross sections are measured as a function of the average transverse momentum, the rapidity separation, and the boost of the two leading jets. The unfolded cross sections agree with perturbative QCD calculations at NLO accuracy apart from phase space regions containing strongly boosted dijets events, in which the measurement is sensitive to the PDFs. Constraints on the PDFs are derived by including the data in a PDF fit together with DIS cross sections from the HERA experiments. Compared to a fit with HERA DIS data alone, the uncertainties of the PDFs, especially those of the gluon PDF, are significantly reduced and a harder gluon PDF is obtained.

Contents

CHAPTER 1

Introduction

The field of particle physics seeks to establish a profound understanding of the fundamental constituents of matter and the interactions between them. The analysis of the scattering products produced in particle collisions at high energies is an excellent method to gain deep insights into the fundamental principles of nature.

In the endeavor to reach the highest energies in order to study very rare particles and to search for physics beyond our current knowledge, particle accelerators have become ever bigger and more complex. Today's most powerful collider is the Large Hadron Collider (LHC), where protons are accelerated to unprecedented energies.

Protons are composed of quarks and gluons, collectively referred to as partons. The structure of the proton is described by parton distribution functions (PDFs), which give the probability to find a quark or gluon at an energy scale Q with a fractional momentum x of the proton. In contrast to the dependence on Q, the x-dependence is not predicted by QCD but has to be parametrized and determined from fits to experimental data. The PDFs are well constrained in the medium-x region from existing measurements but exhibit a large uncertainty in the high-x region which has not yet been accessible by experiments.

When the protons are brought to collision, the partons interact and produce a plethora of new particles. These are detected and measured precisely in large particle detectors installed around the interaction points, such as the Compact Muon Solenoid (CMS) detector. Quarks and gluons, which are also produced abundantly in these collisions, manifest themselves as streams of collimated particles in the detector, the so-called particle jets. The measurement of events containing two such jets with large transverse momenta, known as dijet events, allows for thorough tests of predictions of Quantum Chromodynamics (QCD). The jets are measured precisely in the CMS detector and are compared to perturbative QCD predictions at next-to-leading order accuracy. Ultimately, insights into the proton structure may be gained by confronting predictions with experiments and deriving constraints on the PDFs.

In this thesis, the first measurement of triple-differential dijet cross sections is performed at the LHC. The cross sections are measured as a function of the average transverse momentum, the rapidity separation, and the overall boost of the dijet pair. This triple-differential measurement separates the phase space regions that are sensitive to the PDFs

from those that are not. Fig. 1.1 illustrates the dijet event topologies which are measured in this analysis:

- The dijet measurement is most precise in the **central region** of the detector. For this region, the PDFs are well known and the detector is best understood. The highest transverse momenta of the dijets are reached. Thus, it is very well suited for the determination of the strong coupling constant at high energies.
- The **boosted region** is the most interesting for PDF studies. For dijets with large transverse momenta, the highest x of the proton PDFs become accessible. Since the PDFs are not yet well known in this region, constraints on the PDFs can be derived from a precise measurement.
- Dijet events with a **large rapidity separation** are measured in the same (forward) region of the detector as the boosted dijets. Here, however, the probed medium-x region of the PDFs is already well known. By considering both phase space regions in a PDF fit and exploiting the correlations between them, PDF and detector effects can be disentangled.

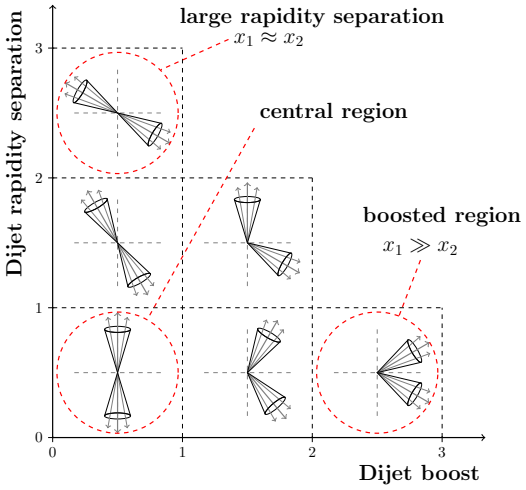

Figure 1.1: Illustration of dijet event topologies measured in the different phase space bins. The most precise dijet measurement is in the central region of the detector. For PDF studies, the boosted region is the most interesting, since the highest x of the proton are accessed. By exploiting the correlations between the bins with dijet events that exhibit a large rapidity separation and the ones with a large boost, PDF and detector effects can be disentangled.

This thesis is structured as follows: In Chapter 2, the theoretical foundations for dijet production at hadron colliders are outlined. An overview of the Standard Model of particle physics with focus on perturbative QCD is given. Furthermore, the relativistic kinematics of dijet production are explained. Chapter 3 summarizes the experimental setup of the CMS detector and the measurement and reconstruction of jets.

The definition of the observables as well as the NLO calculations are discussed in Chapter 4. The measurement of the triple-differential dijet cross section is presented in Chapter 5. A multitude of detector and reconstruction effects are carefully studied and related uncertainties are determined. The measured cross sections, corrected for detector effects in an iterative unfolding procedure, are compared to perturbative QCD calculations at NLO accuracy. The analysis is finalized in Chapter 6 with studies of the PDFs. Constraints on the proton PDFs are presented. In addition, a simultaneous fit of the PDFs and the strong coupling constant is performed.

CHAPTER 2

Theoretical Foundations

A deeper understanding of physical principles always results from the interplay of experimental measurements and the corresponding theoretical predictions. Theoretical models attempt to describe nature's behavior and must be excluded or tentatively confirmed in precise experiments. In the second half of the 20^{th} century, theoretical and experimental physicists have made great progress in describing the fundamental particles and their interactions in a self-consistent model which is today established as the Standard Model of particle physics (SM).

This chapter shortly summarizes the Standard Model while concentrating on quantum chromodynamics (QCD) and those properties, which are the theoretical foundations of this thesis. More extensive and profound discussions can be found in [1–4].

Throughout this thesis, the common unit convention in particle physics is used. It is based on SI units, but is supplemented with the units electron volt (eV) for energy and barn (b) for the interaction cross section. Furthermore, the speed of light c and the reduced Planck constant $\hbar$ are set to unity,

$$c = \hbar = 1.$$

2.1 Standard Model of Particle Physics

The Standard Model of particle physics is a comprehensive theory which describes the fundamental particles and their interactions from first principles. The SM is founded on the concept of quantum field theories, in which the elementary particles and the interactions between them are described by quantized gauge fields. There are four fundamental forces of which three are considered in the SM. While the predictions of the SM have been demonstrated to be extremely robust in a huge variety of experiments, it falls short of being a complete theory of everything, as it does not include a description of the gravitational force and can describe neither dark matter nor non-zero neutrino masses resulting from neutrino oscillations.

Each of the fundamental spin-½ particles, also known as fermions, has a corresponding antiparticle with the same properties but opposite-sign quantum numbers. They are classified into three families and carry quantum numbers of electric charge Q, weak

isospin T and color. The weak hypercharge Y_W is related to the weak isospin and the electric charge by $Y_W = 2(Q - T_3)$.

The electromagnetic and weak forces are described by a U(1)×SU(2) symmetry which is spontaneously broken by the coupling to the scalar Higgs field. The gauge bosons of the unified electroweak theory are a mixture of the gauge bosons of the unbroken symmetry resulting in the massive $W^{\pm}$ and Z^0 bosons and the massless photon. The eigenstates of the weak interaction differ from the mass eigenstates and can be calculated by rotating the mass eigenstates using the CKM matrix [5, 6]. The same effect is observed in the lepton sector in which the mass eigenstates of the neutrinos do not match the interaction eigenstates leading to oscillations between neutrino flavors. The analogous matrix is called PMNS matrix [7, 8].

The strong force is described by the unbroken SU(3) color gauge theory [9, 10], called quantum chromodynamics. The eight gauge bosons of the theory, called gluons, carry color charge.

The Higgs boson, the field quantum of the Higgs field responsible for electroweak symmetry breaking, had long been postulated by theoretical models until its recent discovery at the LHC [11, 12].

2.2 Quantum Chromodynamics

Quantum chromodynamics is the gauge field theory that describes the strong interaction of quarks and gluons. The gauge group of the theory is the special unitary group SU(3). The QCD Lagrangian is given by

$$\mathcal{L} = \sum_q \bar{\psi}_{q,a} \left(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t^C_{ab} \mathcal{A}^C_\mu - m_q \delta_{ab} \right) \psi_{q,b} - \frac{1}{4} F^A_{\mu\nu} F^{A\mu\nu},$$

where γ_μ are the Dirac γ-matrices and $\psi_{q,a}$ are quark-field spinors for a quark with flavor q, mass m_q and a color of index a which runs from 1 to $N_C = 3$. The t^C_{ab} correspond to the eight 3×3 matrices and are the generators of the SU(3) group. $\mathcal{A}^C_\mu$ denotes the gluon fields and C runs from 1 to $N_C - 1$, resulting in eight different gluons. The QCD coupling strength is defined by g_s. The field tensor $F^A_{\mu\nu}$ is given by

$$F^A_{\mu\nu} = \partial_\mu \mathcal{A}^A_\nu - \partial_\nu \mathcal{A}^A_\mu - g_s f_{ABC} \mathcal{A}^B_\mu \mathcal{A}^C_\nu,$$

where f_{ABC} are the structure constants of the SU(3) group. The last term of the field tensor originates from the non-abelian structure, where the generators do not commute, but obey the relation

$$\left[t^A, t^B \right] = i f_{ABC} t^C.$$

This leads to the self-coupling of gluons, one of the prominent features of QCD, resulting in three and four gluon vertex interactions. The fundamental interaction vertices of QCD can be illustrated by the Feynman diagrams in Fig. 2.1. The complete QCD Lagrangian

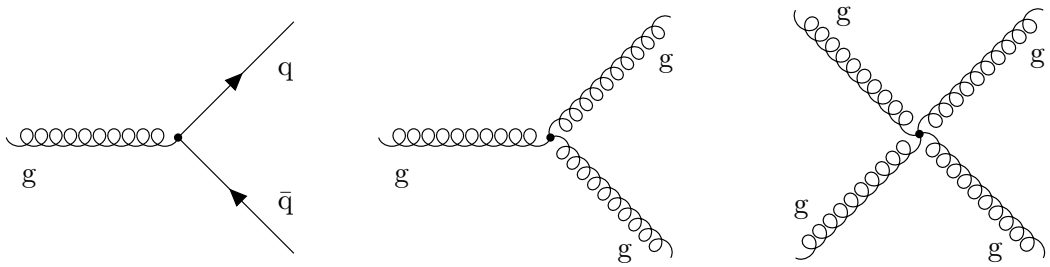

Figure 2.1: The fundamental Feynman rules of QCD comprise a quark-antiquark-gluon vertex, a three-gluon vertex and a four-gluon vertex. The first two are proportional to g_s, the last one is proportional to g_s^2.

also includes a gauge-fixing term and the Faddeev-Popov ghost fields [13], which are not further discussed in this thesis.

Unlike the other fundamental forces, the strong force exhibits two unique and opposing features. Experiments show that quarks and gluons behave like free particles at high energies or small distances. This is called *asymptotic freedom* and is described in QCD by the decreasing of the strong coupling at high energies. At low energies however, quarks and gluons cannot be observed as free particles at larger distances. The strong coupling increases with distance and the creation of new quark-antiquark pairs from the vacuum is energetically favored at some point. This phenomenon, called *confinement*, implies that the strong coupling increases at low energies and is divergent. Consequently, perturbative QCD is not applicable in this energy regime.

2.2.1 Perturbative QCD

All observables X in perturbative QCD (pQCD) are developed as a perturbative series in powers of the strong coupling α_s:

$$X = \sum_{n=0}^{N} \alpha_s^n c_i = c_0 + \alpha_s^1 c_1 + \alpha_s^2 c_2 + \dots,$$

where c_i are the perturbative coefficients. The expansion already yields sufficiently accurate results after the first orders of the perturbative series if $\alpha_s \ll 1$ so that the series converges quickly. However, several features complicate perturbative calculations. Ultraviolet (UV) divergences enter the calculations beyond leading order due to loop corrections. The divergences can be removed by a procedure called renormalization which is described in Sec. 2.2.2. Soft and collinear divergences also need to be handled in perturbative calculations. They arise from singularities at phase-space boundaries and neglected quark masses. Observables must be defined in a way that they are infrared safe and short-distance effects need to be separated from the divergent long-range part, which can be absorbed into the PDFs in a procedure called *collinear factorization*.

2.2.2 Renormalization and Running of the Strong Coupling

Beyond leading order, the calculations include loop corrections which result in UV divergences when calculating the momenta integrals of the loops. To make the result finite, a renormalization procedure is applied. It introduces the renormalization scale μ_r. There are different renormalization schemes, of which the $\overline{\mathrm{MS}}$ scheme [14, 15] is the most popular. Consequently, the observable X and the strong coupling become functions of μ_r.

Nonetheless, the observable X may not depend on the arbitrarily chosen μ_r. The renormalization group equation (RGE) states that the dependence of X on μ_r must cancel. This can be mathematically expressed by

$$\mu_\mathrm{r}^2 \frac{d}{d\mu_\mathrm{r}^2} X\left(\frac{Q^2}{\mu_\mathrm{r}^2}, \alpha_s(\mu_\mathrm{r}^2)\right) = \left(\mu_\mathrm{r}^2 \frac{\partial X}{\partial \mu_\mathrm{r}^2} + \mu_\mathrm{r}^2 \frac{\partial \alpha_s(\mu_\mathrm{r}^2)}{\partial \mu_\mathrm{r}^2} \frac{\partial X}{\partial \alpha_s(\mu_\mathrm{r}^2)}\right) \stackrel{!}{=} 0$$

and states that any dependence of X on μ_r must be canceled by the μ_r-dependence of α_s. Thus, the strong coupling has to fulfill the equation

$$\mu_\mathrm{r}^2 \frac{d\alpha_s}{d\mu_\mathrm{r}^2} = \beta(\alpha_s) = -\left(\beta_0 \alpha_s^2 + \beta_1 \alpha_s^3 + \beta_2 \alpha_s^4 + \ldots\right) \tag{2.1}$$

where β_0, β_1 and β_2 are the 1-loop, 2-loop and 3-loop β-function coefficients, which encode the dependence of the coupling on the energy scale. They are given for the coupling of an effective theory, in which the n_f quark flavors are light ($m_q \ll \mu_\mathrm{r}$). Here, they are given in the $\overline{\mathrm{MS}}$ scheme:

$$\beta_0 = \frac{33 - 2n_f}{12\pi},$$

$$\beta_1 = \frac{153 - 19n_f}{24\pi^2},$$

$$\beta_2 = \frac{2857 - (5033/9)\, n_f + (335/27)\, n_f^2}{128\pi^3}.$$

By integrating Eq. 2.1, the energy dependence of α_s is yielded. Working in an energy range of constant number of flavors i. e. if no quark mass thresholds are passed, an analytic solution exists in 1-loop approximation:

$$\alpha_s(\mu_\mathrm{r}^2) = \frac{1}{\beta_0 \ln\left(\mu_\mathrm{r}^2/\Lambda^2\right)}$$

Λ is the constant of integration and corresponds to the scale at which the perturbative coupling would become large and the perturbative series diverge. It is very often convenient to give the strong coupling at a specific scale, from which the coupling at any scale is calculated. It is common practice to report $\alpha_s(M_\mathrm{Z})$, the strong coupling at the scale of the Z boson mass. Thus, the 1-loop analytical function can be expressed as

$$\alpha_s\left(\mu_\mathrm{r}, \alpha_s(M_\mathrm{Z})\right) = \frac{\alpha_s(M_\mathrm{Z})}{1 + \alpha_s(M_\mathrm{Z})\beta_0 \ln\left(\mu_\mathrm{r}^2/M_\mathrm{Z}^2\right)}.$$

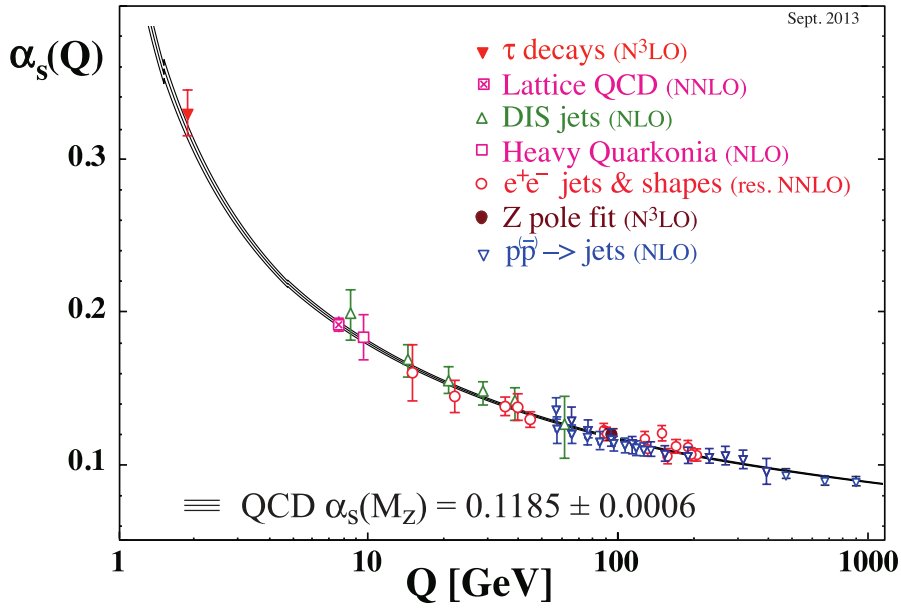

Figure 2.2: Running of the strong coupling constant as predicted by QCD. Determinations of the strong coupling from several experiments are shown at the scale of the measurement and confirm the running up to 1 TeV. Recent measurements from CMS [16, 17] probe the running at even higher scales. Figure taken from [2].

As the parameter α_s is a free parameter of the theory, it is deduced from experimental measurements and evolved to the scale of the Z-boson. The current world average value of the strong coupling according to the PDG [2] reads as

$$\alpha_s(M_Z) = 0.1181 \pm 0.0013$$

and is determined from hadronic τ lepton decays, lattice QCD calculations, deep inelastic scattering data, electron-positron annihilation processes and electroweak precision fits. Fig. 2.2 shows various determinations of the strong coupling from measurements at scales Q, which describe the running of the strong coupling up to the 1 TeV scale.

2.2.3 Factorization and Parton Density Functions

The *collinear factorization* allows to separate the calculation of an observable into a short-distance part, calculable in QCD, and an approximately but universal long-distance part, which is described by parton distribution functions. The factorization analogously to the renormalization involves an arbitrary choice of a factorization scale μ_f. Particle emissions with transverse momenta above μ_f are included in the hard scattering perturbative coefficients while emissions softer than μ_f are accounted for within the PDFs.

Parton Distribution Functions The structure of the proton is described by parton distribution functions (PDFs) in which the partons represent the constituents of the

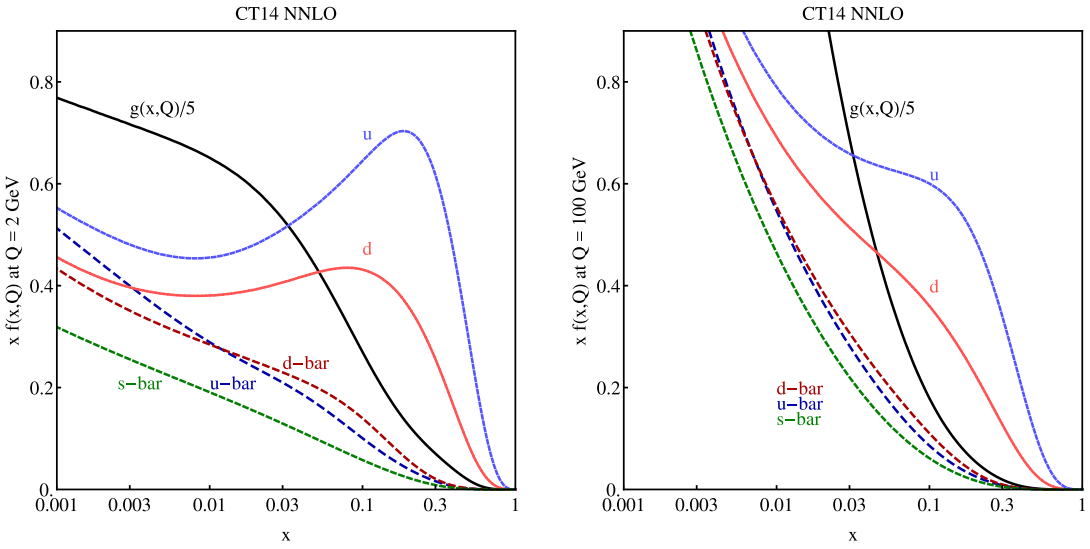

Figure 2.3: The CT14 NNLO PDF set with the gluon and quark PDFs at the scale $Q = 2\,\text{GeV}$ (left) and evolved to $Q = 100\,\text{GeV}$ (right) using the DGLAP evolution equations. Figure taken from [19].

proton. The PDFs give the probability density[1] to find a parton carrying a momentum fraction x of the proton momentum at an squared energy scale Q^2. As only the Q-dependence is predicted by QCD, the x-dependence needs to be parametrized and determined in fits to experimental data.

Several groups obtain the proton PDFs in global fits to a large variety of measurements from different experiments. Instead of determining all (13 quark, antiquark and gluon) PDFs independently, the number can be reduced by choosing a sufficiently low starting scale Q_0 below the threshold of the charm quark mass and calculating the heavy flavor PDFs in a heavy flavor scheme [18]. Furthermore, the PDFs of the top and anti-top quark are often neglected due to their large rest mass. In a fit, each parton is parametrized with a sufficiently flexible function to describe the x-dependence at the starting scale Q_0. The PDFs are then evolved to the scales of each measurement and the PDF parameters are adapted in an iterative least-squares fit.

Global PDFs are determined by the CTEQ [19], MMHT [20], NNPDF [21] and the ABM [22] groups at LO, NLO and NNLO. While the range of measurements which are put into the fit are often similar, there are differences in the applied minimization method, the phenomenological approaches and the estimation of the uncertainties. More details are given in Sec. 4.4.2, in which the uncertainty of the PDFs on the cross section measurement is discussed.

The HERAPDF group [23] uses a slightly different approach by using a less flexible parametrization but restricting the data to measurements from the HERA experiments, which provide a very precise and compatible data set. Furthermore, the HERAPDF group

[1] More specifically a number density, as the PDFs are normalized to the number of partons.

made their fitting framework XFitter [24] freely available as open-source software. XFitter is employed in Sec. 6 of this thesis to study the constraints on the PDFs provided by the triple differential dijet measurement.

DGLAP Evolution Equations Because of the factorization, the PDFs depend on the scale μ_f. This evolution from a scale Q_0 to a different scale Q can be calculated using the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) [25–27] equations, which have a different structure for the gluon and quark PDFs. Using the shorthand notation

$$[f \otimes P] = [P \otimes f] = \int_x^1 \frac{d\xi}{\xi} f\left(\xi, \mu_\mathrm{f}^2\right) P\left(\frac{x}{\xi}\right)$$

they can be expressed at leading order for the quark PDFs $q_i(x, \mu_\mathrm{f})$ as:

$$\frac{\partial q_i(x, \mu_\mathrm{f}^2)}{\partial \ln(\mu_\mathrm{f}^2)} = \frac{\alpha_s(\mu_\mathrm{f}^2)}{2\pi}\Big([q_i \otimes P_{qq}] + [g \otimes P_{qg}] \Big),$$

and for the gluon PDF $g(x, \mu_\mathrm{f})$ as

$$\frac{\partial g(x, \mu_\mathrm{f}^2)}{\partial \ln(\mu_\mathrm{f}^2)} = \frac{\alpha_s(\mu_\mathrm{f}^2)}{2\pi}\left(\left[\sum_i q_i \otimes P_{gq}\right] + [g \otimes P_{gg}]\right).$$

where the sum runs over all quark and antiquark flavors. P_{ab} are the so-called Altarelli-Parisi kernels, also known as splitting functions. These give the probability to emit a parton a with momentum fraction x from a parton b with momentum fraction ξ.

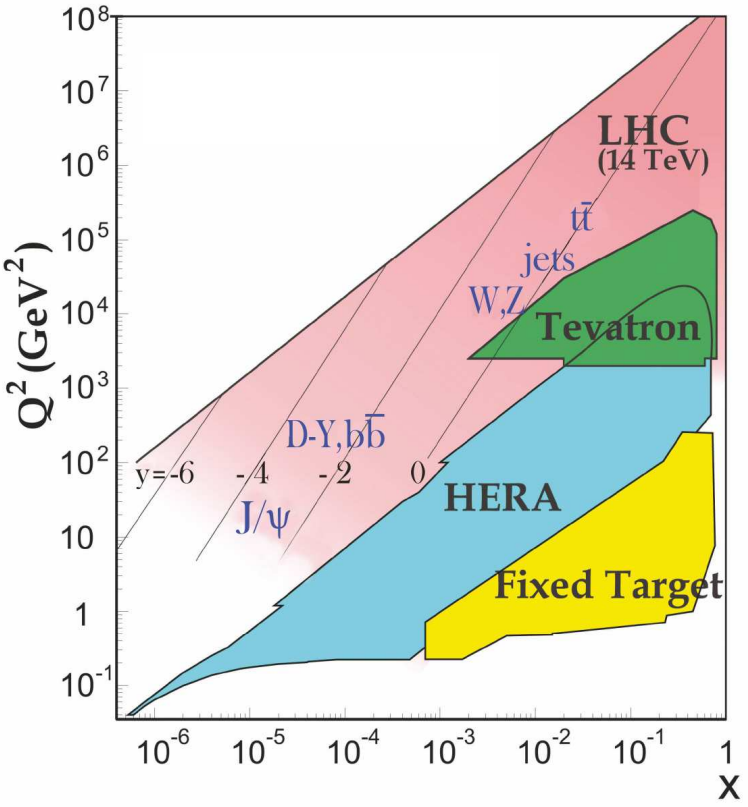

Figure 2.4: The phase space in x and Q^2 which is accessible in current experiments. The most important input for the PDFs continues to be provided by DIS data measured at the HERA collider. Further important constraints at high x and low scales are provided by fixed-target experiments. The latest PDF sets also include LHC measurements which provide constraints at high energy scales Q^2 and high x (jets, $t\bar{t}$) and at medium x (W,Z,+jets). Figure taken from [2].

2.3 Hadronization and Parton Shower

As perturbative QCD calculations lack the capability of describing the soft component of the interaction, additional models are employed which describe the emission of additional partons and the hadronization into colorless bound states.

2.3.1 Parton Shower

Following the hard scattering process, the accelerated colored partons undergo subsequent emission of gluons. Unlike for QED radiation, gluons themselves carry color charge and therefore also emit further gluons, leading to a shower of colored partons, called the parton shower.

The dominant contributions of the parton shower come from collinear parton splitting and soft gluon emissions. The collinear splitting of a parton is described by splitting func-

tions which are identical to the DGLAP splitting functions. The successive application of the splitting to the colored particles leads to the parton shower. The evolution of the shower is determined by an evolution variable. Common choices for this variable involve the virtual mass square of the partons in the shower, the transverse momentum or the angle. The parton shower is terminated when the scale is below the hadronization scale, which is of order 1 GeV.

In actuality, the parton shower mimics the effect of higher-order corrections. As it is often not feasible to calculate these, the parton shower approximation is used instead, although great care has to be taken to avoid any double counting if using a NLO generator in combination with a parton shower.

2.3.2 Hadronization

The result of the parton shower is a large number of color charged particles. As objects with color charge cannot be observed as free particles, they have to hadronize into bound states which are colorless. The MC event generators Pythia and Herwig employ different phenomenological models to simulate the hadronization process. Pythia uses the Lund string fragmentation model while Herwig is based on the cluster fragmentation model.

Lund String Fragmentation Model Within the Lund string model, the attraction between a $q\bar{q}$ pair is modeled using so-called strings, whose energy follows a Coulomb potential as a function of the distance between the two quarks [28]. Final-state gluons from the parton shower are considered as kinks in the strings. If the string exceeds a certain energy threshold, it breaks up and new quark-antiquark pairs are formed. If the available energy is too small, the quarks and antiquarks recombine into mesons and baryons.

Cluster Fragmentation Model First, all gluons are split into quark antiquark pairs. Neighboring pairs are grouped together to form colorless clusters [29, 30]. Most of these clusters decay into hadrons in an isotropic two-body phase space model.

2.4 Jets and Jet Algorithms

Particle jets are the clustered streams of particles which were produced by the hadronization of quarks or gluons. Since they provide the link between the short-scale physics and the observed final states, they yield important information about the PDFs and the hard interaction.

There are many different algorithms available which cluster jets from a set of input objects. Both CMS and ATLAS rely on the anti-k_{T} and inclusive-k_{T} jet algorithms, which have proved to be very robust and are both collinear and infrared safe, see Sec. 2.4.1. They are sequential recombination algorithms and combine input objects based on a distance measure in Minkowski space. All jets in this thesis were reconstructed using the efficient algorithms implemented in the FastJet library [31].

2.4.1 Collinear and Infrared Safety

Hard partons undergo many collinear splittings during the fragmentation process. Furthermore, there are always emissions of soft particles in QCD-like events caused by non-perturbative and perturbative effects. If the set of hard jets in an events remains unchanged by those effects, they are considered to be infrared and collinear safe [32].

Fig. 2.5 shows the effect of a collinear splitting (right plot) and of additional soft particles (left plot) on the results of an unsafe jet algorithm. Most of the cone-based jet algorithms which cluster elements using a constant distance measure in the η-ϕ space are affected by the previously mentioned issues. This explains the popularity of the modern sequential recombination algorithms which are used in almost all of today's jet-based analyses.

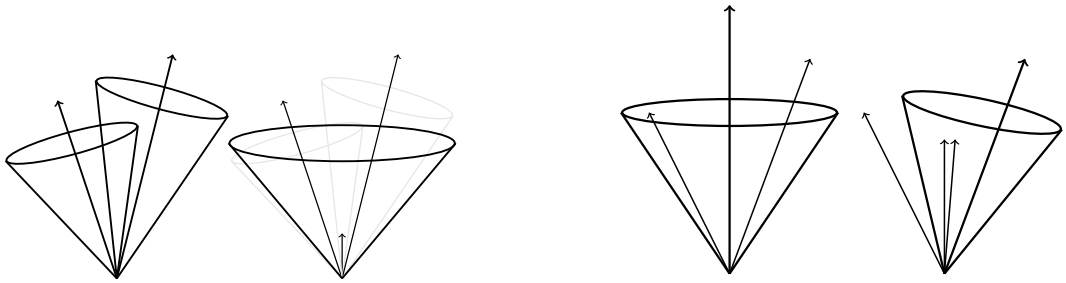

Figure 2.5: The influence of an infrared emission and a quasi collinear splitting on a jet algorithm not fulfilling the infrared safe and collinear safe requirements is shown. An additional soft particle (left plot) leads to the combination of the two jets into one large jet. A quasi-collinear splitting of a particle (right plot) changes the result of the jet clustering algorithm.

2.4.2 Generalized k_{T} Jet Algorithms

The most popular sequential recombination algorithms are the k_{T} jet algorithms. They cluster jets based on a jet size parameter R and an additional parameter p which introduces a dependence on the transverse momentum of the input objects. The pairwise algorithm uses a list of input objects which can be partons, stable particles or reconstructed particle candidates.

First, the distance d_{ij} between two particles i and j and the distances d_{iB} and d_{jB} of the particles to the beam are calculated based on the rapidity difference Δy_{ij} and the azimuthal angle $\Delta\phi_{ij}$ between them:

$$d_{ij} = \min(p_{\mathrm{T}i}^{2p}, p_{\mathrm{T}j}^{2p}) \frac{(\Delta R_{ij})^2}{R^2}$$
$$d_{i\mathrm{B}} = p_{\mathrm{T}i}^{2p}$$

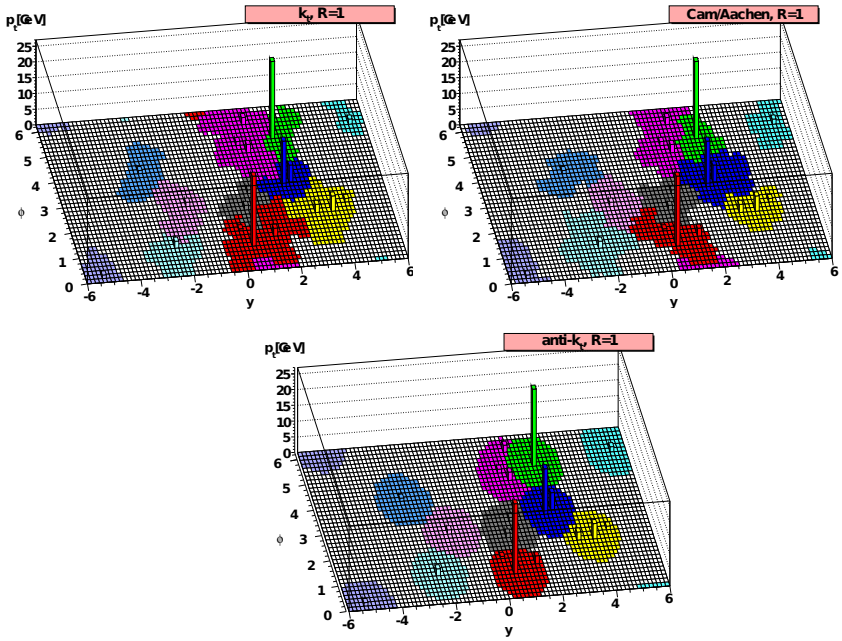

Figure 2.6: Catchment areas of jets obtained by the described k_T-based algorithm. For most jet algorithms the shape is irregular, while the anti-k_T algorithm yields circular shapes for hard jets while and crescent-shaped soft jets. Adapted from [32].

with the angular distance

$$(\Delta R_{ij})^2 = (\Delta y_{ij})^2 + (\Delta \phi_{ij})^2$$

If the distance d_{ij} is smaller than the distances to the beam line, the two particles i and j are merged into a new particle k which then replaces the particles i and j in the input list. These steps are repeated until all particles are clustered into jets. Since the distance measures are defined in Minkowski space, the shapes of the jets in the η-ϕ plane are not circular but irregular, see Fig. 2.6. Based on the parameter p, there are three important k_T based algorithms with distinct properties:

- $p = 1$: The **Inclusive $\mathbf{k_T}$ algorithm** [33, 34] is based on a p_T^2 distance measure and approximately describes the inversion of the QCD branching process.
- $p = 0$: The **Cambridge-Aachen algorithm** [35] is only based on the spatial separation of the objects and does not rely on the transverse energy of the input objects. Similarly to the inclusive k_T algorithm, it produces jets of irregular shape. While it is not widely used for jet analyses, it is very interesting in the context of jet substructure studies, where at first a jet with a large jet size is clustered and subsequently its structure is investigated.
- $p = -1$: The **anti-$\mathbf{k_T}$ algorithm** [36] favors clustering hard input objects resulting in fairly circular jet shapes for hard jets, while soft jets in close proximity to a

reconstructed hard jet are crescent-shaped. It is the most widely used jet algorithm for physics analyses at the LHC.

2.5 Dijet Production at Proton Colliders

When two protons collide, the actual interaction takes place between the constituents of the proton, the partons. The cross section σ of such a hard scattering process can be expressed in collinear factorization as

$$\sigma_{P_1P_2\to X} = \sum_{i,j} \int dx_1 dx_2 f_{i,P_1}\left(x_1, \mu_\mathrm{f}^2\right) f_{j,P_2}\left(x_2, \mu_\mathrm{f}^2\right) \times \hat{\sigma}_{i,j\to X}\left(x_1 p_1, x_2 p_2, \alpha_s(\mu_\mathrm{r}^2), \frac{Q^2}{\mu_\mathrm{f}^2}\right),$$

where f_i and f_j denote the parton distributions, which depend on the momentum fraction x of the parent proton P and the factorization scale μ_f. The parton-level cross section $\hat{\sigma}$ for the production of the final state X depends on the final state phase, the factorization scale and the renormalization scale. The sum runs over all contributing initial state partons. Fig. 2.7 illustrates the factorization into the PDFs and the hard scattering cross section.

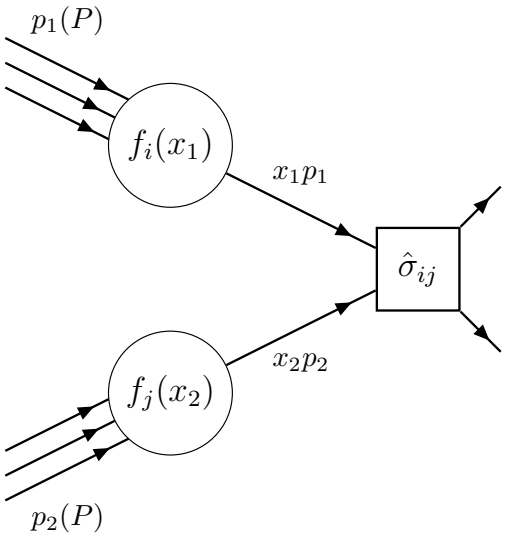

Figure 2.7: The total cross section is factorized into the hard scattering cross section $\hat{\sigma}$ and the PDFs $f_i(x)$.

2.5.1 Kinematics of the Dijet System

The outgoing partons of a hard interaction manifest themselves as streams of collimated particles which are clustered into jets. Consequently, the properties of jets are studied in order to gain a deeper understanding of QCD and the proton structure. In the following discussion, the incoming partons involved in the scattering process are assumed to be massless and collinear to the beam protons. Furthermore, it is convenient to describe the kinematics of the dijet system using the transverse momentum p_{T} and the rapidity y of the jets, as it is introduced in Sec. 3.3.1. The rapidity y, defined as

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

is at lowest order directly related to the proton momentum fractions x_1 and x_2 via

$$x_1 = \frac{x_{\mathrm{T}}}{2} \left(e^{y_1} + e^{y_2} \right) \qquad \text{and} \qquad x_2 = \frac{x_{\mathrm{T}}}{2} \left(e^{-y_1} + e^{-y_2} \right),$$

with $x_{\mathrm{T}} = p_{\mathrm{T}}/E$ and the rapidities y_1 and y_2 of the two outgoing partons.

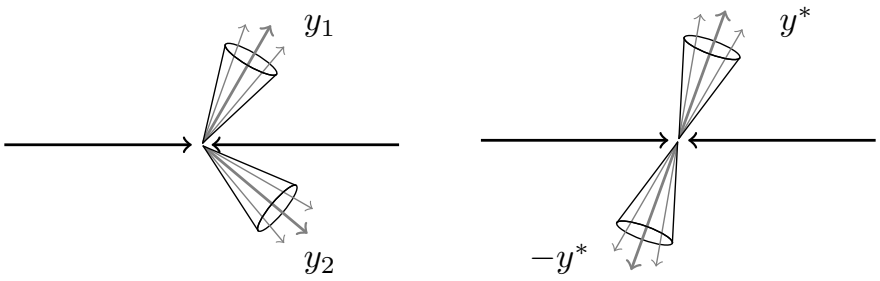

Figure 2.8: Dijet event in the laboratory frame (left) with the rapidities y_1 and y_2 of the jets and in the center-of-mass frame (right) with the rapidities $\pm y^*$.

The longitudinal boost of the parton-parton center-of-mass (CM) frame with respect to the proton-proton CM frame, y_{b}, is calculated from the rapidities y_1 and y_2 of the two jets emerging from the partons.

$$y_{\mathrm{b}} = \frac{1}{2} |y_1 + y_2|$$

In the center-of-mass (CM) frame, the rapidities of the jets can be expressed using the variable y^*. As y^* is symmetric, it is defined as

$$y^* = \frac{1}{2} |y_1 - y_2|.$$

The quantity y^* may also be expressed in terms of the polar scattering angle θ^* with respect to the beamline by

$$y^* = \frac{1}{2} \ln \left(\frac{1 + |\cos \theta^*|}{1 - |\cos \theta^*|} \right).$$

CHAPTER 3

Experimental Setup

The European Organization for Nuclear Research (CERN) was founded in 1954. Originally dedicated to the study of atomic nuclei, it is now devoted to the research of sub-atomic particles and their interactions. To accomplish that task, CERN built several particle accelerators reaching record-breaking energies and explored energy ranges which had not been accessible before.

Ground-breaking achievements like the discovery of neutral currents in the Gargamelle bubble chamber [37], the discovery of the W and Z boson by the UA1 and UA2 experiments [38] at the SPS accelerator or the creation of antihydrogen atoms were accomplished by physicists at CERN.

Remarkable insights into the Standard Model were gained by measurements at subsequent collider experiments. The mass of the Z and W boson were precisely measured at the LEP collider. The proton-antiproton collider Tevatron at FNAL discovered the top quark and measured its mass accurately. Since the search for the long anticipated Higgs boson was unsuccessful due to their limited energy reach, an even larger and more powerful accelerator was planned and built, the Large Hadron Collider (LHC). The search for the Higgs boson finally succeeded in 2012 [11, 12], but many further questions like the nature of dark matter or the existence of supersymmetry are yet to be resolved.

3.1 The Large Hadron Collider

The LHC is the world's most powerful particle accelerator and collider. It is contained in the circular tunnel of the preceding LEP collider which has a circumference of 27 km at a depth between 50 m and 175 m. The tunnel crosses the border between Switzerland and France at four points and two of the four main experiments are located in France.

Two adjacent beamlines that intersect at four interaction points contain the particle beams travelling in opposite directions. More than 1000 dipole magnets generating a magnetic field of up to 8.3 T bend the beams on a circular track while almost 400 quadrupole magnets keep the beams focused.

The beams are brought to collision at four interaction points which house the LHC experiments ALICE [39], ATLAS [40], CMS [41–43] and LHCb [44]. ALICE is designed to study the quark-gluon plasma produced by colliding heavy ions, which resembles the

initial state of the universe. LHCb is precisely measuring the CP violation and the decay of B mesons. ATLAS and CMS, general purpose detectors which allow a broad field of physics studies, were built to search and study the Higgs boson and physics models beyond the Standard Model. Furthermore, precision measurements of Standard Model predictions and its parameters improve the current knowledge and confidence in its predictions.

Prior to injection and acceleration of protons in the LHC, the particles pass a series of consecutive acceleration steps, successively increasing their energy. The linear particle accelerator (LINAC2) generates 50 MeV protons that are further accelerated in the Proton Synchrotron Booster (PSB) and the Proton-Synchrotron (PS) to 26 GeV. The Super-Proton-Synchrotron (SPS) further accelerates the protons up to an energy of 450 GeV. At last, the proton beams are injected into the LHC ring in which they are accelerated up to peak design energy of 13 TeV. All these pre-accelerators are not only used to feed the LHC, but also serve other physics experiments like the radioactive ion beam facility ISOLDE, see Fig. 3.1.

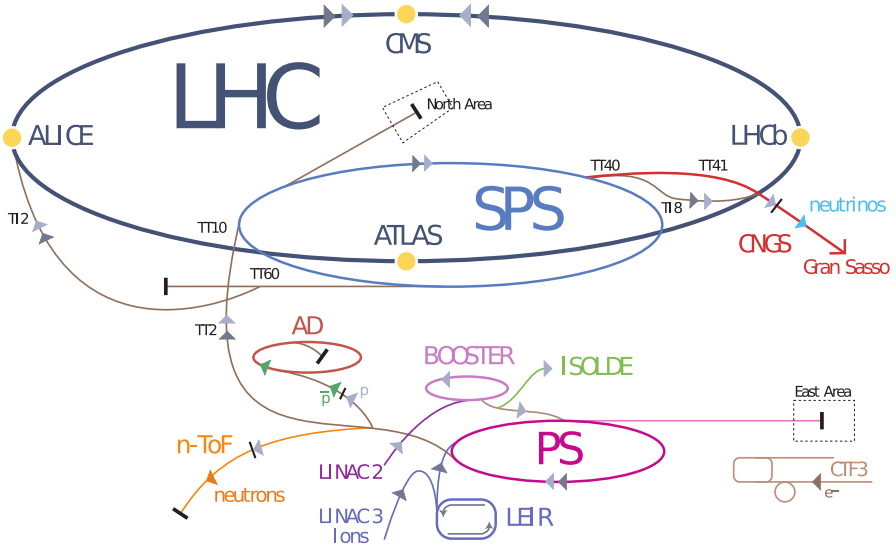

Figure 3.1: Several particle accelerators are chained together to feed proton beams into the LHC. Further experiments are located along the accelerator complex serving a broad program of physics studies [45].

3.2 Luminosity measurement

The cross section σ of a physical process is related to the event rate $\dot{N}$ by the luminosity $\mathcal{L}$,

$$\dot{N} = \mathcal{L}\sigma.$$

The luminosity is dependent on the particle beam parameters and can be expressed by

$$\mathcal{L} = \frac{n_p^2 n_b f_{\text{rev}} \gamma F}{4\pi \epsilon_n \beta^*},$$

where n_p is the number of particles per bunch, n_b is the number of bunches per beam, f_{rev} the revolution frequency, γ the relativistic gamma factor and F the geometric luminosity reduction factor. The effective collision area of the two beams is related to the normalized transverse beam emittance ϵ_n and the value of the betatron function β^* at the interaction point.

The instantaneous luminosity is constantly monitored by the experiments. CMS employs two methods to estimate the relative instantaneous luminosity [46]. The first method measures the particle flux in the hadron forward calorimeter which is related to the instantaneous luminosity. The second counts the number of clusters in the pixel tracking detector measured in zero-bias events. The absolute luminosity measurement is relying on van-der-Meer scans carried out in special runs of the LHC [47]. The luminosity measurement is affected by an uncertainty, which propagates on any absolute cross section measurement, see Sec. 5.8.1.

Fig. 3.2 shows the integrated luminosity delivered by the LHC to the CMS experiment in the run periods from 2010 to 2012.

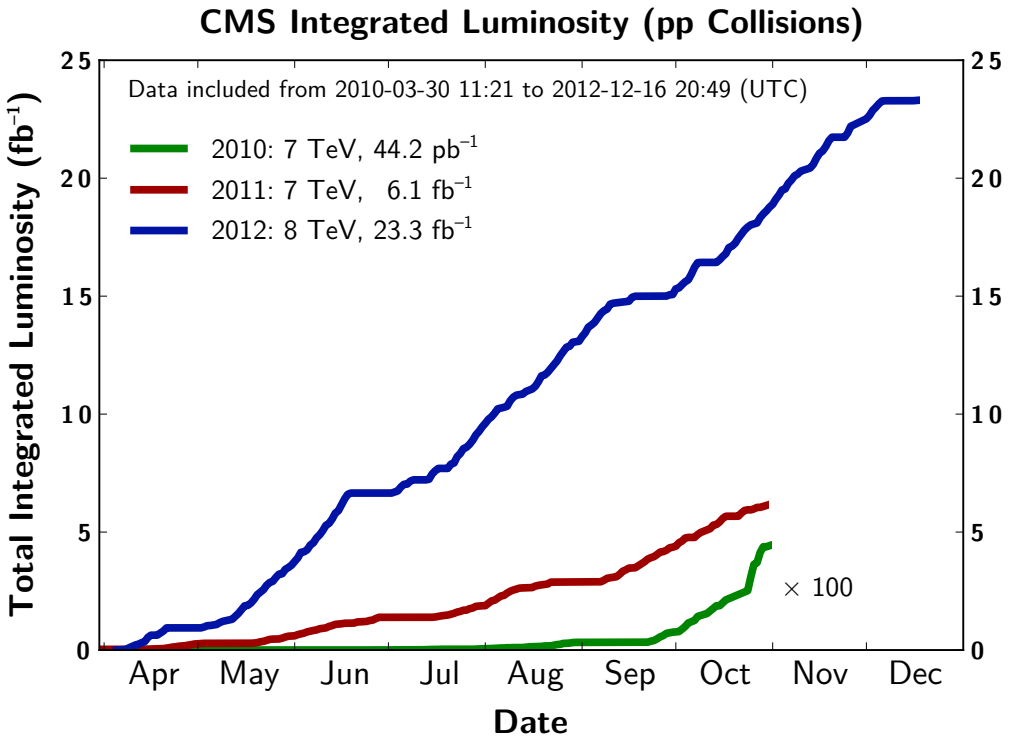

Figure 3.2: Integrated luminosity delivered by the LHC to CMS in the 2010, 2011 and 2012 LHC run periods. Taken from [48].

3.3 The Compact Muon Solenoid Detector

The Compact Muon Solenoid (CMS) detector is a general purpose detector at the LHC, located at Point 5 of the LHC ring. To serve a wide range of physics studies, the detector design is driven by a cylinder-shaped structure containing layers of different subdetectors, each built to measure a specific type of particles with best precision, see Fig. 3.3. A high-precision inner tracking system is surrounded by an electromagnetic and a hadronic calorimeter which again are enclosed by a superconducting solenoid magnet. The whole inner part of the detector is surrounded by a sophisticated muon detection system embedded in an iron yoke.

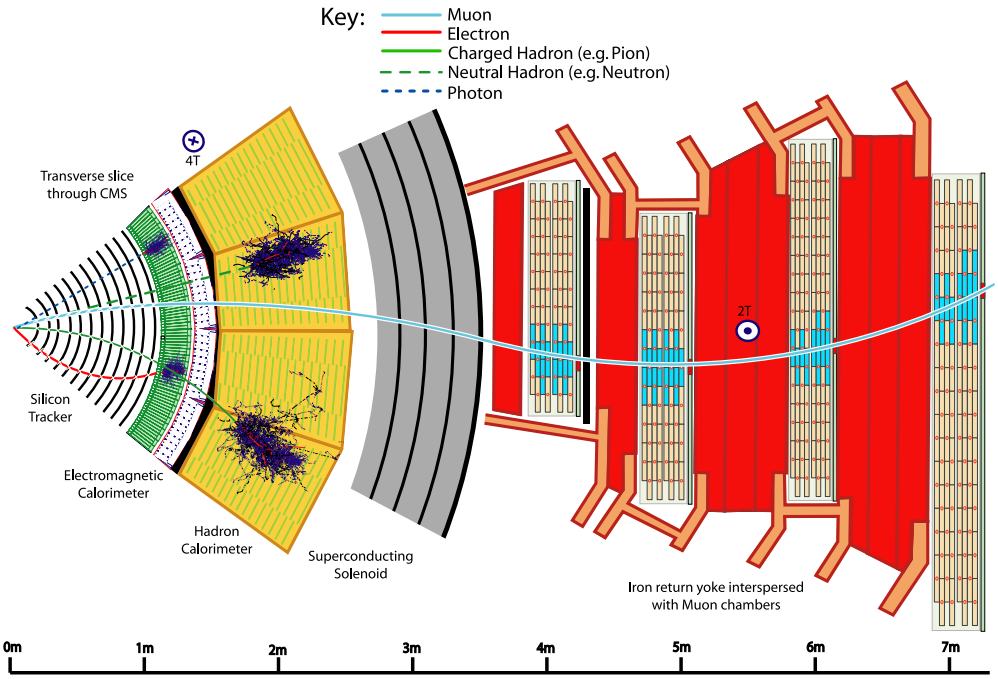

Figure 3.3: A transverse slice through the CMS detector shows the various subdetectors and the signatures by which passing particles are detected. Taken from [49].

The detector is 21.6 m long and 14.6 m in diameter, but weighs more than 12 000 t due to its compact design. It was constructed as cylindrical slices constructed at ground level and lowered into the cavern. In case of upgrades or repairs, the slices can be pulled apart and the inner components can be easily accessed. A longitudinal section of one quadrant of the CMS detector, which reveals the location and coverage of all subdetectors, is shown in Fig. 3.4.

The detector and physics performance of the CMS detector are discussed in great detail in [41–43]. This section intends to only present a short overview of the design and functional principles of the detector.

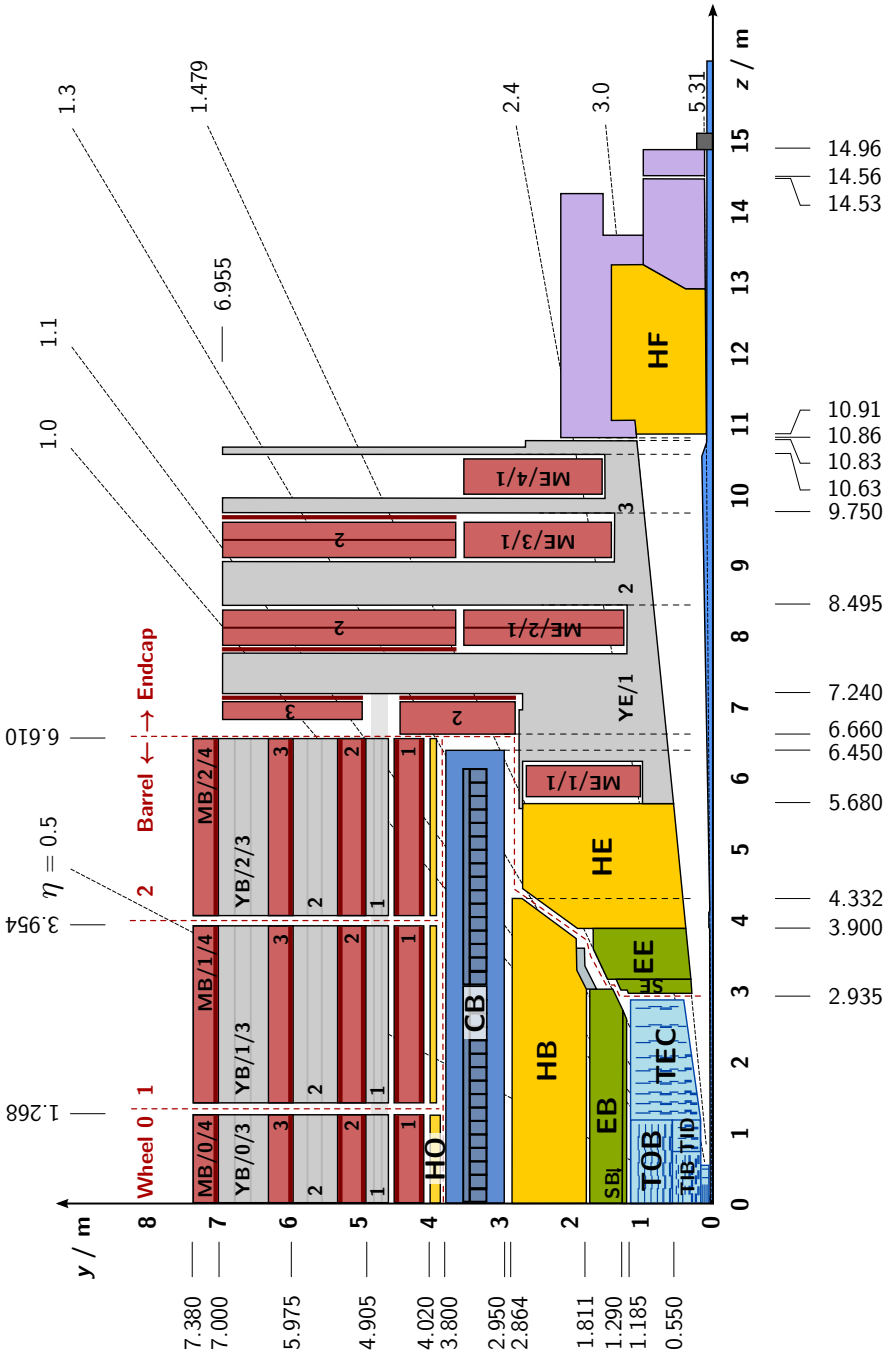

Figure 3.4: A longitudinal section of one quadrant of the CMS detector in the y-z plane. The sketch shows the multi-layer design of the CMS detector starting with the silicon pixel and silicon strip detectors close to the interaction point. They are surrounded by the electromagnetic (green) and hadronic (yellow) calorimeters. The barrel is encompassed by the superconducting magnet (blue). The muon detection system (red) is embedded in the iron return yoke. Taken from [48].

3.3.1 Definition of the Coordinate System

CMS uses a right-handed coordinate system centered at the nominal interaction point inside the detector. The x-axis points horizontally towards the center of the LHC ring neglecting the small tilt of the LHC ring, the y-axis vertically upwards and the z-axis along the beam direction towards the Jura mountains. Important quantities are the azimuthal angle ϕ, measured from the x-axis in the x-y plane, and the polar angle θ, measured from the z-axis in the z-y plane. Instead of the polar angle θ, the pseudorapidity η and the rapidity y are commonly used to divide the phase space. The pseudorapidity is defined as

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right).$$

Throughout this thesis, the rapidity is favored over the pseudorapidity. Rapidity differences are invariant under longitudinal boosts, which does not hold for the pseudorapidity. Rapidity and pseudorapidity are equivalent in case of massless particles. The rapidity is defined as

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

The momentum along the beamline is not well-defined due to the momentum distribution inside the proton. A direct connection to the hard process is given by the transverse momentum p_T related to Cartesian coordinates as

$$p_\mathrm{T} = \sqrt{p_x^2 + p_y^2}\,.$$

3.3.2 Inner Tracking System

In order to yield a best possible spatial resolution, the particle tracks need to be measured as close to the beam line as possible. The inner tracking system of CMS consists of silicon detectors which measure the hits of charged particles emerging from the collision.

The silicon detectors are depleted from free charged by applying a voltage. When charged particles pass through the detector material, they leave a small ionization current which can be detected and measured as a hit in the detector. By combining multiple hits, the track of a charged particle can be reconstructed and the momentum and charge of the particle can be determined based on a mass hypothesis. Due to the strong magnetic field of the CMS detector, even tracks of particles with high transverse momenta have a measurable curvature.

The inner tracking detector encloses the interaction point with a diameter of 2.6 m and extends up to 2.8 m in each direction along the beamline, the tracking system covers a pseudorapidity range up to $|\eta| < 2.4$. The inner tracking system comprises two subsystems, the silicon pixel detector consisting of three layers which is installed very close to the beam pipe and the silicon strip detector located further outside with ten strip layers in the barrel region, see Fig. 3.5.

Silicon Pixel Detector Containing over 65 million pixels arranged in three cylindrical layers at 4 cm, 7 cm and 11 cm distance to the beam pipe, the pixel detector is able to resolve the tracks of a the huge number of particles. At LHC design luminosity, about 1000 particles pass the tracking detector on average per bunch crossing. The size of each pixel is 100 µm x 150 µm giving an average occupancy of 10^{-4} per bunch crossing. The high spatial resolution achieved by the pixel detector furthermore allows the identification and measurement of secondary vertices used to identify long-lived particles.

Silicon Strip Detector The pixel detector is complemented by a silicon strip detector. Reduced particle flux further away from the beam pipe eases the identification of tracks. Cost-efficient silicon strips are employed reaching out to a radius of 1.3 m. The strip detector consists of a total of 10 million detecting strips which are read out by 80 000 chips. To avoid any blind detector area, the strips are arranged overlapping.

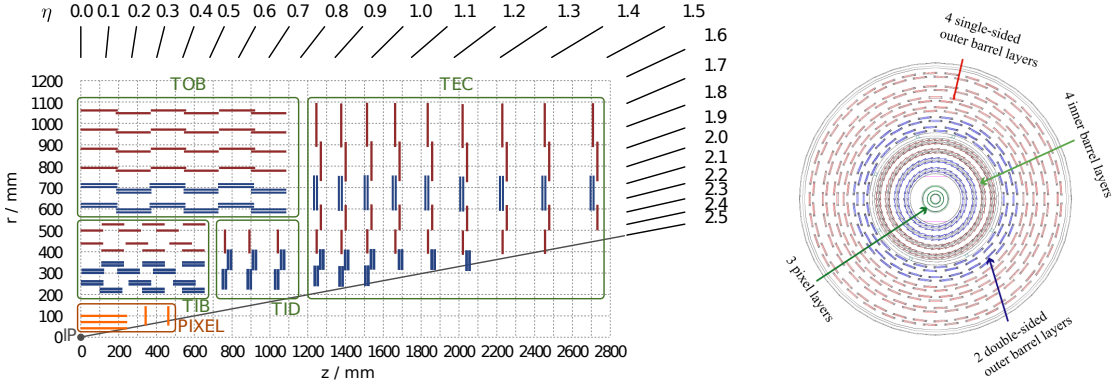

Figure 3.5: The left plot shows one quadrant of a longitudinal section of the inner tracking system consisting of the silicon pixel detector and the silicon strip detector. The right figure shows a transverse section of the tracking system in the barrel region which nicely illustrates the overlapping arrangement of the strip modules. Figures taken from [48] and [50], respectively.

3.3.3 Electromagnetic Calorimeter

Measuring only the tracks of traversing particles is not sufficient to identify the particles and derive their momentum. The energy needs to be determined as well by stopping the particles in the detector and summing up the deposited energy. The photon and electron energy is measured in the electromagnetic calorimeter (ECAL).

High-energy photons, electrons or positrons which enter the dense material of the ECAL detector produce an electromagnetic shower via subsequent bremsstrahlung and electron-pair production processes. Below a certain threshold, the particles deposit their energy via Compton scattering and the photoelectric effect in the detector material resulting in an excitation of the material atomic state. Subsequently, they emit photons which are measured using avalanche photodiodes. The fraction of the deposited energy is proportional to the number of emitted photons.

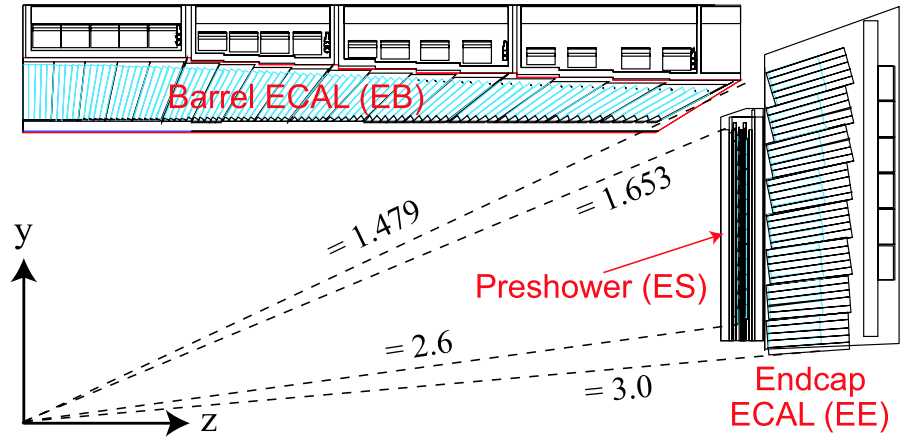

Figure 3.6: The electromagnetic calorimeter consists of submodules covering the barrel region (EB) and the endcaps (EE). A complementary preshower detector (ES) is mounted in front of the endcaps. Taken from [41].

The hermetic calorimeter is made of lead tungstate ($PbWO_4$), a very dense material with a radiation length of $X_0 = 0.89\,\text{cm}$. Because of the incorporated oxygen, it is highly transparent and scintillates light. The small Molière radius of 2.19 cm leads to a fine granularity. These material properties allow the ECAL to be built very compact and to be placed within the solenoid magnet.

Figure 3.6 shows a schematic sketch of the ECAL in the y-z plane. The ECAL comprises three subsystems covering the pseudorapidity range up to 3.0.

Electromagnetic Calorimeter Barrel (EB) The EB extends up to $\eta < 1.479$ using more than 60 000 crystals which form a homogeneous coverage in pseudorapidity. Each crystal measures $2.2\,\text{cm} \times 2.2\,\text{cm} \times 23\,\text{cm}$ which corresponds to 25.8 X_0 radiation lengths.

Electromagnetic Calorimeter Endcaps(EE) The ECAL endcaps seal off the barrel region and extend the pseudo rapidity coverage in the region $1.479 < |\eta| < 3.0$ with an additional 15 000 crystals.

Electromagnetic Pre-shower Detector (ES) To increase the spatial precision, the EE is complemented with the ES, which sits in front of it and consists of two orthogonal silicon strip sensors. The ES improves the discrimination between single high-energy photons and less interesting low-energy photon pairs as well as the discrimination between neutral pions and photons.

The relative energy resolution of the ECAL can be parametrized using the NSC-formula

$$\left(\frac{\sigma}{E}\right)^2 = \frac{N^2}{E^2} + \frac{S^2}{E} + C^2,$$

in which the first term describes the contribution by noise (N), the second term the

stochastic (S) component arising from the proportional relation between the number of counted photons and the deposited energy, and last a constant (C) offset term.

3.3.4 Hadronic Calorimeter

Hadrons entering the calorimeter produce hadronic showers. High-energy hadrons mostly shower in inelastic interactions producing a large number of pions and nucleons. Due to the large transverse momentum of the secondary particles, hadronic showers spread further in the calorimeter than electromagnetic showers. When the energy of the particles involved in the shower drops below a certain threshold, the energy is deposited by ionization and low-energy hadronic activity. The active scintillation material is excited and emits blue-violet light. All scintillators are connected to photodiodes using wavelength shifters which read out the signals and pass them to the data acquisition system.

The compact design of the CMS detector limits the size of the calorimeters. CMS therefore built a sampling calorimeter inside the solenoid coil. The hadronic calorimeter consists of brass as absorber material because it is non-magnetic and has a short interaction length of $\lambda_I = 16\,\text{cm}$. It is interleaved with plastic scintillators measuring the deposited energy. The CMS hadronic calorimeter comprises three subsystems.

Hadron Barrel Calorimeter (HB) It covers the barrel region up to a pseudorapidity $|\eta| < 1.305$. The absorbing material in the barrel has a corresponding thickness of $5.39\,\lambda$ in the central region and up to $10.3\,\lambda$ at $|\eta| = 1.3$. The HB is complemented by the Hadron Outer Calorimeter (HO) located on top of the coil of the magnet. Using the coil as absorbing material, it is able to measure the tails of hadron showers penetrating the HB and the coil.

Hadron Endcap Calorimeter (HE) The HE extends the pseudorapidity coverage up to $|\eta| < 3.0$. A major challenge in the construction of the HE were the usage of non-magnetic material in order to not disturb the magnetic field as well as the close distance to the beamline. Radiation damages decrease the detector response which has to be corrected continuously.

Hadron Forward Calorimeter (HF) The forward calorimeter extends even closer to the beam pipe. With a coverage of $2.8 < |\eta| < 5.2$ the calorimeter is adapted to the high radiation environment. The HF is built using iron absorbers and quartz fibers as active material, which measure the Cerenkov light emitted by the relativistic components of the shower.

3.3.5 Superconducting Solenoid

A key component of the CMS detector is the superconducting magnet which produces a magnetic field with a strength of 4 T and is located inside the detector between the calorimeters and the muon system. It measures a diameter of 6 m and a length of 12.5 m. When operated at design magnetic field strength, the magnet contains an energy of 2.6 GJ.

The strong magnetic field is necessary to bend the tracks of particles with high momentum to achieve a good resolution in the tracking system. Operated at a temperature of 4 K, the NbTi conductors become superconducting. The magnet is complemented by a 10 000 t iron yoke which returns the magnetic field.

3.3.6 Muon System

Identifying and measuring muons with high precision is an unrivaled capacity of the CMS detector. Unlike most other particles, muons are not stopped by the calorimeters but leave the detector. Therefore, the muon system has been placed around the other detector components in the iron return yoke of the magnet to measure the bent tracks of the muons.

By combining the information of the inner tracking system and the muon detectors, the path and the muon momentum are both measured precisely. The muon system comprises three different types of detectors each suited for a specific task. Drift tubes (DT) cover the barrel region up to $|\eta| < 1.2$, the endcaps up to $|\eta| < 2.4$ contain cathode strip chambers (CSC) which also work reliably in the spatially varying magnetic field. The DT and CSC detectors yield a precise spatial muon resolution. Both systems are accompanied by resistive plate chambers (RPC) which provide fast response to the trigger system.

3.3.7 Trigger and Data Acquisition

The LHC generates a huge number of collisions. At beam crossing frequencies of 25 ns, there are 40 million bunch crossings per second with an average of around 20 collisions per bunch crossing in the 2012 run period. With today's hardware, the storage of all collision events is not feasible. Furthermore, most of the collisions are soft and of low interest for physics analyses. Therefore, a complex trigger system consisting of a very fast component implemented in hardware, the Level 1 trigger (L1) and a High Level software trigger (HLT) analyze the events and accept only events which are interesting for physics analyses.

L1 Trigger At the same frequency as collisions occur, the L1 trigger reads out the detector electronics and analyzes the data using custom hardware. The workflow of the L1 trigger is shown in Fig. 3.7 left. Trigger Primitive Generators calculate the transverse energy and missing energy from the front-end electronics readout. Regional Calorimeter Triggers (RCT) identify electromagnetic showers in the ECAL and sum up ECAL and HCAL trigger towers. Furthermore, pattern recognition is performed to identify jets and hadronic τ decays. A jet candidate is found, if the transverse energy in a region of 4×4 trigger towers is greater as or equal to the transverse energy of the eight surrounding regions, see Fig. 3.7 right. These candidates are passed to the Global Calorimeter Trigger (GCT), which sorts the incoming candidates from all 18 regional triggers and passes the top candidates to the Global Trigger (GT). The GT accepts events with a frequency of 100 kHz and passes them to the data acquisition system, which processes the data and transfers them to the HLT.

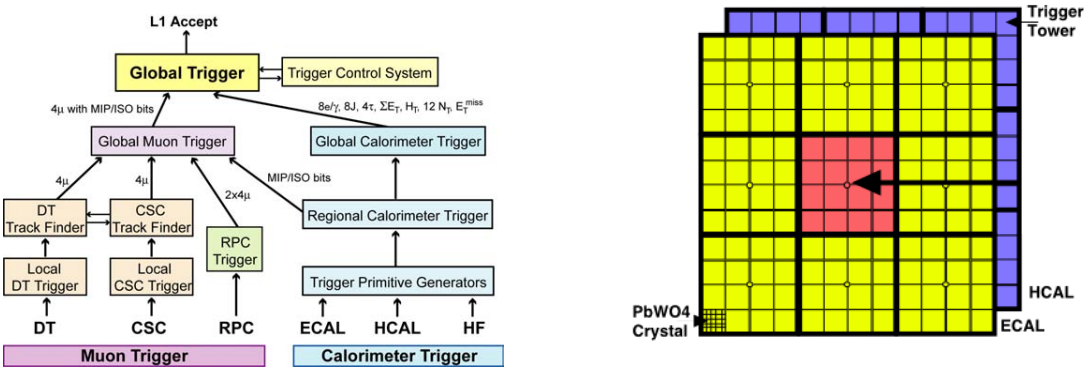

Figure 3.7: Left: Workflow of the L1 trigger system. The regional triggers search for jets and compute the transverse and missing energy of an event. The global calorimeter trigger sorts the objects from the regional calorimeter triggers and passes the top candidates to the Global trigger, which accepts or rejects the event. If it is accepted, the complete data and the trigger objects are passed to the DAQ. Right: Jet candidates in the Level 1 calorimeter trigger are formed from 4×4 trigger towers. Figures taken from [41] and [51].

HLT Trigger The HLT is a software-based trigger running on a dedicated computing farm at Point 5. The software is implemented in a streamlined version of the CMS software framework. Each HLT path is a sequence of reconstruction and selection steps with increasing complexity. In the end, the HLT accepts several 100 events per second for permanent storage and analysis.

Jets are reconstructed in the HLT using the anti-k_{T} jet clustering algorithm. Because of the high processing time of the Particle Flow algorithm, see Sec. 3.5.1, the jet trigger paths are divided into multiple selection steps. At first, jets are reconstructed from calorimeter towers. Only for events in which at least one calorimeter jet passes a certain p_{T} threshold, the Particle Flow algorithm is run and the jets are clustered again from the Particle Flow candidates. Due to the flexibility of the HLT, it is already possible to apply sophisticated jet energy corrections during the HLT selection.

Data Acquisition (DAQ) As the L1 trigger accepts events at a rate of 100 kHz, the DAQ system has to process the events at the same speed. It reads out the data of all detector subcomponents and assembles the complete events, see Fig. 3.8. The data is subsequently passed to the HLT which further reduces the rates to a few hundred events per second. Finally, the events are merged and saved to a local storage system, from which they are continuously transferred to the Tier-0 computing center at CERN.

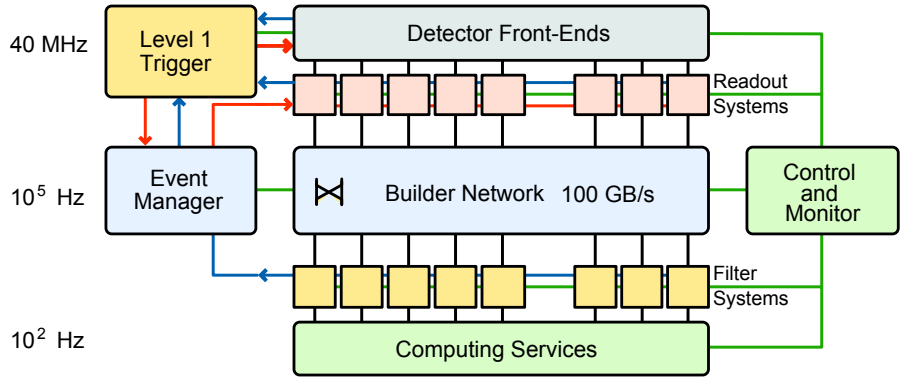

Figure 3.8: The L1 trigger accepts events at a rate of 100 kHz and passes them to the DAQ. The DAQ reads out all detector signals, builds the complete event and passes it to the HLT. All events accepted by the HLT are stored and transferred to the Tier-0 computing center. Taken from [41].

3.4 Computing Infrastructure and Software Tools

The vast amount of data produced at the LHC experiments and the complexity of the software pose many challenges for the computing infrastructure and the engaged software. On the theory side, powerful Monte Carlo event generators, which are able to simulate the collision events need to be developed. On experimental side, the physical event information needs to be reconstructed from the raw detector readout. Furthermore, the complex architecture of the detector response needs to be modeled and simulated. These central tasks are approached using a common software framework within CMS, the CMSSW framework, which interfaces the various theory tools and all the reconstruction and detector software. All arising processing tasks are divided into smaller units, called jobs, which are assigned to computing centers distributed over many countries. This common computing and storage infrastructure is called the worldwide LHC computing grid (LHCG).

3.4.1 Worldwide LHC Computing Grid

To overcome the discussed challenges and to ease the access of users to the data of the LHC experiments, a distributed grid with a tiered infrastructure was developed. As the majority of the data is produced at CERN, a hierarchical structure with the computing center at CERN at the top was chosen, see Fig. 3.9. The raw data is stored at CERN and distributed to globally distributed Tier-1 centers, as they provide further storage resources and large computing resources for the reconstruction and analysis of the data. Tier-2 sites provide additional computing resources while Tier-3 sites are mostly used by local groups for data analysis.

Access to the resources of the LHCG is gained by certificates which authorize the user to access the storage and computing resources.

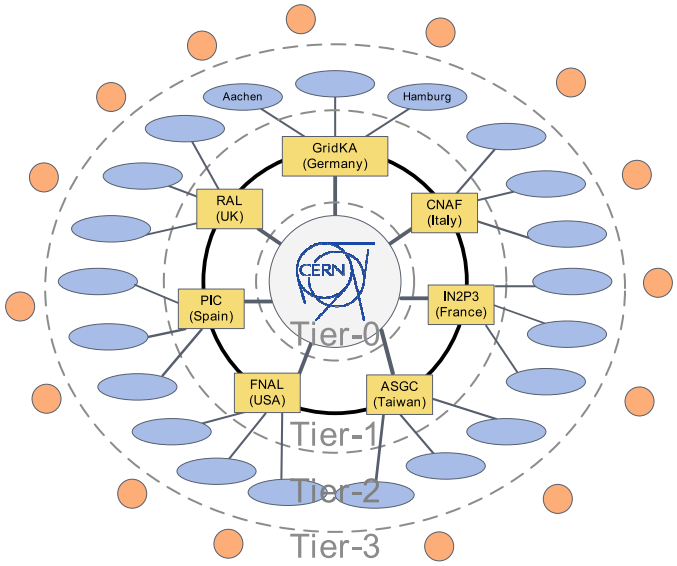

Figure 3.9: The Worldwide LHC Computing Grid is ordered hierarchically with the CERN Tier-T0 at the top. Taken from [52].

3.4.2 Dynamic Cloud Computing Resources

The static computing resources available in the WLCG are often not optimally utilized if used for cpu intensive tasks. It is much more efficient to dynamically allocate virtual machines in a shared HPC cluster when they are needed and shut them down afterwards. The NLO cross section calculations for the triple-differential dijet cross sections are extremely cpu intensive but involve only small input/output data transfers. Thus, they are perfectly suited for calculations on an remote HPC cluster such as the bwForCluster in Freiburg, which was integrated into the local computing infrastructure of the institute during the Master's thesis of [53].

3.4.3 CMS Software Framework

The software framework of the CMS collaboration (CMSSW) [54], offers all necessary tools for a physics analysis. The tasks in the event processing comprise on the one hand calibration and reconstruction of data from raw detector read-out and on the other hand the event generation and detector simulation. Furthermore, it provides the possibility to implement analysis code to perform the data analysis.

To cope with this vast range of requirements to the experiment software, CMSSW is built on top of an event data model (EDM), in which the event is a container for all measured or simulated data. The reconstruction and distribution algorithms in CMSSW are divided into modules, which can be dynamically loaded and run. Each module reads the event data and can add additional objects to the event. The execution of modules is ordered in processing chains which can be configured by the user. Very often these modules access external libraries like Monte Carlo event generators for event simulation,

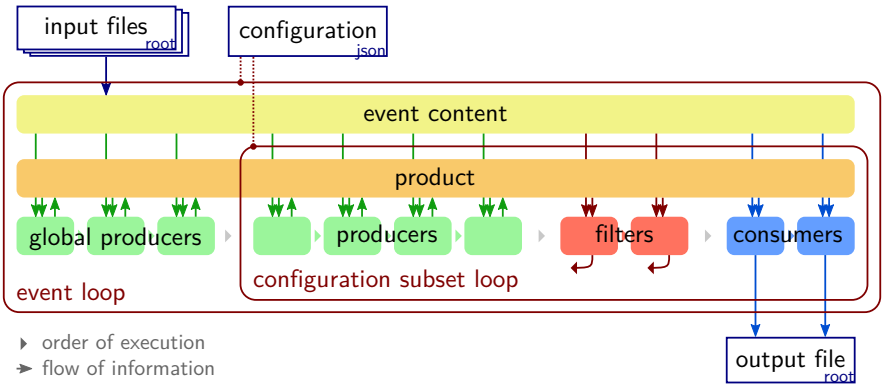

Figure 3.10: Workflow of an analysis in the Artus framework. Taken from [48].

Geant 4 for the detector simulation or FastJet for the reconstruction of jets.

While having that much information available in the event data is convenient to redo reconstruction steps, it is unsuited for the fast processing of the analysis due to its size and complexity. Therefore a skimming step, in which only the necessary data is preserved, is run before the analysis, see Sec. 3.4.4.

3.4.4 Analysis Software and Workflow

Due to the complexity of the workflows in the HEP data analysis, several analysis tools were used or even developed in the Karlsruhe group to facilitate a reliable and fast workflow of the analysis.

Artus and Kappa

The Kappa software [55] is a skimming framework interfaced to CMSSW. It consists of different modules which allow to skim only the physics objects needed in the subsequent analysis. The data is stored in its distinct compact data format using the ROOT object serialization capabilities. The resulting Kappa tuples provide all necessary information of the events and the lumisections, while hiding the complexity of the CMSSW data sets.

The analysis itself is built on top of the Artus framework [56]. Artus has been developed within the Karlsruhe group to combine analysis efforts and to profit from mutual developments. The framework defines a workflow based on three elements, see Fig. 3.10. There are *producers*, which calculate quantities and *filters*, which reject events based on the defined criteria. In a final step, histograms or tuples are written out by *consumers*. All producers, filters and consumers are written in a modular way so that they can be shared with other analyses. Furthermore they are steered by a global configuration file in which all settings and cuts can be easily adapted.

grid-control

The data sets are even after the skimming step too large to be processed on a single computer. Therefore special batch systems with a large number of computing nodes are employed to process the data. The data processing is split into multiple jobs which are then sent to the batch system. grid-control [57] is by far the most versatile job submission tool which provides multiple options for data splitting and parametrized jobs while hiding the pitfalls of local or remote computing resources.

ROOT

The object-oriented data analysis framework ROOT [58] has been written more than 20 years ago. However it is still very popular and used by all LHC experiments for persistent storage of data. Moreover, ROOT provides fast histogramming classes and access to many libraries like MINUIT for minimization purposes or TMVA for multivariate data analyses. Despite many hours of headaches caused by obscure design decisions in the software, ROOT and especially the python bindings PyROOT are used extensively in this analysis.

matplotlib and numpy

matplotlib is a 2D plotting library written in the python programming language [59]. It provides publication quality plots in a variety of output formats and is very pleasant to use. All plots in this thesis were made using matplotlib. The plotting library matplotlib as well as many scripts used in this thesis rely on the scientific computing library NumPy [60]. NumPy provides powerful n-dimensional arrays and tools to manipulate them.

3.4.5 Monte Carlo Event Generators and Simulation Software

Pythia The multi-purpose event generator Pythia simulates events in high-energy collisions, comprising a large set of physics processes. Pythia uses the Lund string hadronization model in which all but the highest-energy gluons are treated as field lines which attract each other by gluon self-interaction and form a tube or string of strong color field. In this analysis two version of the Pythia event generator are used. The official samples including the detector simulation were generated using Madgraph and Pythia 6 [61], while the study of non-perturbative effects was performed using the new Pythia 8 [62] version, in which all the employed tunes are available.

Herwig Herwig is also a multi purpose event generator for the simulation of high-energy hadron-hadron collisions. The first version was build in Fortran and is known as HERWIG [63]. Herwig++ [64] builds up on the heritage of the HERWIG version while providing a much more flexible structure as it is implemented in C++. The recently released Herwig 7 [65] version combines all their developments and supersedes both version.

The Herwig generator family includes all steps to simulate events. It includes a number of hard scattering processes, but also possesses the possibility to interface external matrix element generators. The parton shower simulates initial- and final-state radiation via angular ordering, multiple partonic scatterings are simulated by an eikonal model and a cluster model describes the hadronization.

Herwig 7 further improves these capabilities by including next-to-leading order QCD matrix elements with matched parton showers while keeping the key features of the previous Herwig versions. Herwig++ is used in this thesis to study non-perturbative effects, see Sec. 4.3. The NLO capabilities of Herwig 7 are shown in the comparison of the unfolded measurement to NLO predictions with matched parton showers, see Sec. 5.9.

POWHEG POWHEG describes a method by which fixed NLO calculations can be merged with parton showers [66, 67]. Furthermore, there is a software called POWHEG box [68], which provides a general framework for the implementation of NLO calculations. It contains the hard matrix elements for NLO dijet production. For the parton shower and hadronization, it needs to be interfaced to a Monte Carlo event generator like Pythia or Herwig.

MadGraph MadGraph [69] is an automated multi-purpose tree-element matrix element generator. It implements a large number of processes and can be interfaced to Monte Carlo Event generators. In this thesis the MadGraph matrix elements are used together with the Pythia 6 event generator for general comparisons to data.

NLOJet++ and FASTNLO The complicated NLO cross sections for jet production are calculated using NLOJET++ [70]. It implements the dipole subtraction method for the separation of the divergences. NLOJET++ can calculate up to three-jet observables at NLO precision. It implements the ability to run user analysis scenarios by which it is interfaced to the FASTNLO project [71, 72].

Since the pQCD cross section calculations in NLOJET++ are determined in Monte Carlo integration and are therefore very time consuming, it is not feasible to repeatedly calculate the cross sections as it is necessary for PDF fits or uncertainty estimations. The FASTNLO framework implements a strategy for fast recalculations of cross sections. It stores the perturbative coefficients obtained with NLOJET++ in a way that the strong coupling constant and the PDFs can be changed afterwards without a recalculation of the perturbative coefficients.

LHAPDF All event generators and cross section calculation tools need the parton distribution functions as input. They are either hard coded in the generator or accessed using a standardized interface, the LHAPDF library [73, 74]. LHAPDF stores the PDFs in a discretized structure in data files. It provides interpolation routines to read the PDFs and interpolate the PDFs at all scales. LHAPDF is used by almost all major MC generators.

3.5 Reconstruction of Jets

In scattering processes with large momentum transfers, the outgoing partons produce a collimated stream of particles when hadronizing. The clusters of these particles are the experimental signatures of quarks and gluons in the detector and are called *jets*. The clustering of the particles is performed using jet algorithms, which are discussed in Sec. 2.4. An important property of jet algorithms is their applicability to all kinds of input objects, i. e. partons, stable particles or reconstructed particle candidates. Consequently, the clustered jets are called parton jets, particle jets, and reconstructed jets, respectively[1]. Fig. 3.11 illustrates the different levels, at which jets can be clustered.

In the CMS detector, jets show up as localized deposit of energy in the calorimeters accompanied by a large number of tracks in the direction of the deposited energy. The particle candidates, input of the jet algorithm, are reconstructed using different techniques based on the amount of information available. If jets are reconstructed from the energy clusters within the calorimeters, they are called *calorimeter jets*. If the jet reconstruction uses Particle Flow candidates, they are called *Particle Flow jets*. By removing pileup tracks from the Particle Flow candidates before the jet clustering, one yields *Particle Flow CHS jets*.

In the analysis which is presented in this thesis, all jets are clustered using the anti-k_{T} algorithm with a jet size parameter of $R = 0.7$. When talking about jets which are clustered at reconstructed level, it is always referred to Particle Flow CHS jets.

3.5.1 The Particle Flow Algorithm

CMS uses the Particle Flow reconstruction algorithm [75, 76] to identify and reconstruct particles by combining information from all detector subsystems. Due to its compact design inside the solenoid, the hadronic calorimeter is not able to stop and measure all particles which reduces the energy resolution. However, taking into account the additional information of the tracking system within the Particle Flow algorithm enhances the reconstruction performance and leads to a jet energy resolution comparable to ATLAS. A key element in the Particle Flow algorithm is the strong magnetic field of CMS which allows the precise distinction between neutral and charged hadrons.

The ingredients of the Particle Flow algorithm are the tracks and vertices, reconstructed from hits in the tracking detectors, the deposited energy in the electromagnetic and hadronic calorimeters and the tracks in the muon system. The reconstructed particles are classified as muons, electrons, photons, charged hadrons or neutral hadrons. The combination of all sub-detectors yields an optimal identification and measurement of their momentum and energy.

The tracks are found using the Combinatorial Track Finder (CTF) algorithm [77] employed by CMS. Based on these tracks, the primary vertices in an event are identified. The electromagnetic and hadronic calorimeters are divided into a grid of cells based on the detector granularity to identify calorimeter cluster seeds. If there are seeds with an

[1] In analogy, the corresponding levels are later referred to as parton level, particle level and reconstructed level.

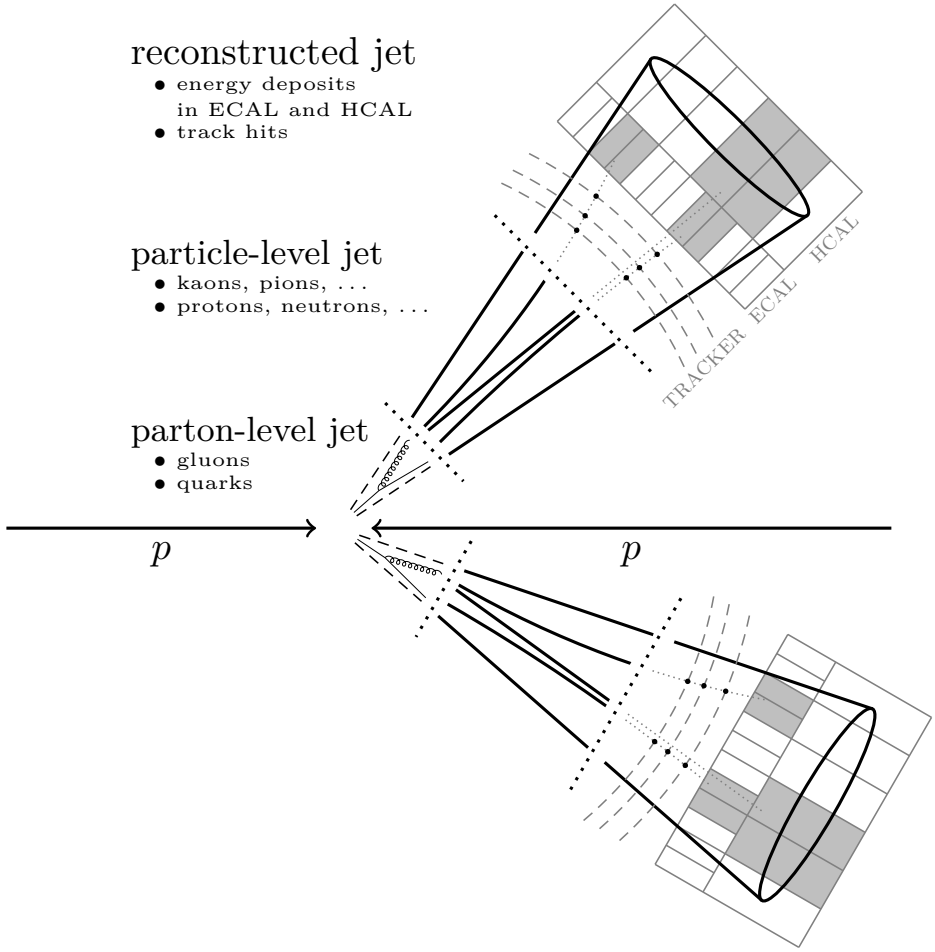

Figure 3.11: Illustration of different levels at which jets can be reconstructed. If partons or stable particles are clustered into jets, parton jets or particle jets are yielded, respectively. Jets clustered from reconstructed particle candidates are called reconstructed jets.

energy exceeding a certain threshold, they are used in an iterative merging algorithm to form Particle Flow clusters. The different elements of the detector information are then linked together into Particle Flow building blocks based on the geometry and χ^2 fits. At first, muons, which can be well identified using the tracks in the muon detector, are reconstructed using the building blocks connected to the muon system. Blocks connecting the inner tracking system with the ECAL clusters are used to identify electrons. Similarly, charged hadrons are identified using links between the tracking system and the remaining calorimeter clusters. Only neutral objects which leave no traces in the tracking system, remain. ECAL clusters are interpreted as photon candidates, while the remaining HCAL clusters are assumed to be deposits of neutral hadrons. To avoid any kind of double-counting of energy, all Particle Flow building blocks of successfully reconstructed particle candidates are removed and the energy of the calorimeter clusters is recalculated.

Finally, a set of so-called Particle Flow candidates is yielded. They consist of well identified particles, which profit from the improved resolution gained by the inclusion of tracking information. This collection of particles is then used to reconstruct the jets and further physical objects.

3.5.2 Jet Area

The jet area describes the space covered by a jet object in the η-ϕ plane [78]. While cone-based jet algorithms yield a πR^2 size area, the area for sequential jet algorithms needs to be determined for each jet. A large number of infinitely soft particles, so-called ghost particles, are evenly distributed in the event. The area A_j of a jet j is assumed to be proportional to the number of ghost particles clustered into the jet.

The concept of jet areas is especially important in the context of pileup mitigation. The average p_T density ρ in an event is estimated using the inclusive k_T algorithm which also clusters many soft jets in the event and covers the entire ϕ-η phase space. The average p_T density is defined as

$$\rho = \mathrm{median}\frac{p_\mathrm{T}^j}{A_j}.$$

ρ is a measure for the underlying event and pileup activity in the event and is used later on to correct the jets for these effects, see Sec. 3.5.4.

3.5.3 Charged Hadron Subtraction

CMS introduced a new technique to reduce pileup from jets using the high resolution of the tracker, called Charged Hadron Subtraction (CHS) [79]. Naturally the CHS algorithm can only be applied on jets within the tracker coverage of $|\eta| < 2.4$. All tracks of Particle Flow candidates which originate from a pileup vertex are removed, see Fig. 3.12. Tracks originating from the signal vertex or tracks not associated with any vertex remain in the event. The signal vertex is defined as the vertex with the largest sum of p_T squared of its associated tracks. The CHS jets are clustered from this remaining set of particles. Since the jet identification criteria applied by the jet selection in the analysis require at least one charged particle in a jet, the CHS method is effectively reducing the influence

of pileup on real jets as well as completely removing a majority of the pileup jets in the barrel region of the CMS detector in combination with the jet identification criteria.

The analysis presented in this thesis relies on CHS jets. Especially in the low-p_T region, removing pileup jets mimicking the leading jet in the event improves the signal efficiency in the forward region.

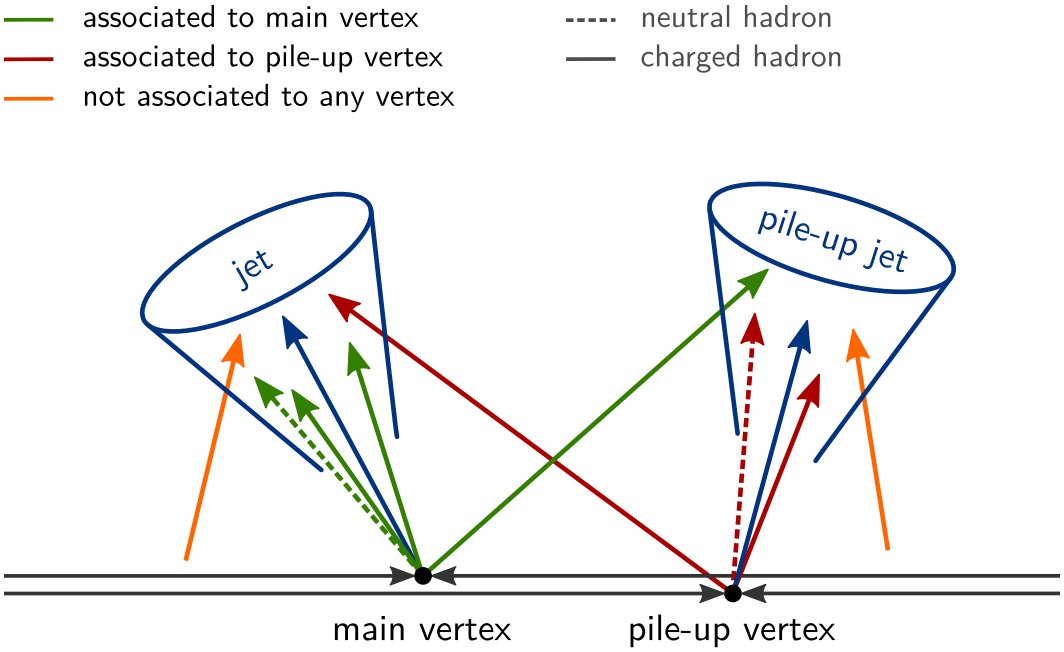

Figure 3.12: Illustration of particles clustered into a jet which originates from the main vertex or into a pileup jet. The CHS algorithm removes charged particles not originating from the main vertex before clustering the jets. Taken from [48].

3.5.4 Jet Energy Corrections

A detailed understanding of the jet energy scale and the jet transverse momentum resolution is important when it comes to drawing conclusions about the properties of quarks and gluons produced in high-energy scattering processes. On the experimental side, there are multiple effects causing the reconstructed jet energy not to correspond to the true jet energy like electronic noise, pileup and underlying event effects, but also non-linearities in the calorimeter response and numerous further small effects. The jet reconstruction and definition itself also introduces effects due to the fragmentation model, initial- and final-state radiation which can cause out-of-cone effects.

The jet energy corrections (JEC) relate the measured jet energy to the corresponding true particle jet energy and are derived using sophisticated methods by the JetMET group [80, 81]. CMS uses a factorized correction approach consisting of multiple correction steps that build on one another. The corrected transverse momentum p_T^{corr} of a jet is

yielded by subsequently applying all correction factors on an uncorrected jet

$$p_{\mathrm{T}}^{\mathrm{corr}} = c_{\mathrm{res}}(\eta, p'_{\mathrm{T}}) \cdot c_{\mathrm{mc}}(\eta, p''_{\mathrm{T}}) \cdot c_{\mathrm{pileup}}(\eta, \rho, A_j, p_{\mathrm{T}}^{\mathrm{raw}}) \cdot p_{\mathrm{T}}^{\mathrm{raw}},$$

where $p_{\mathrm{T}}^{\mathrm{raw}}$ is the transverse momentum of the uncorrected jet, p'_{T} is the transverse momentum after applying the pileup correction factor c_{pileup}. p''_{T} is the transverse momentum after the additional correction c_{MC} of relative and absolute effects derived from MC studies. Finally a correction for residual effects c_{res} derived from data is applied to yield the corrected jet transverse momentum.

Pileup Corrections The first step removes the effects of pileup contamination. Additional soft proton-proton interactions produce particles being clustered into the jets originating from the hard interaction. This additional amount of energy needs to be subtracted by the pileup correction. The applied correction is based on the jet area method by using the pileup density ρ in the event and the jet area A_j. The raw jet energy is then corrected by a factor proportional to the pileup density and the jet area.

MC Corrections Based on simulated QCD events, the jet energy scale is further corrected. The momentum of a reconstructed jet can differ from a generated particle jet due to out-of-cone effects or detector inefficiencies, which are modeled by the detector simulation. The inverse response of the reconstructed jet to a generated jet is applied as correction to remove these effects.

Residual Data Corrections Additional effects which can not be reliably estimated in a Monte Carlo simulation are corrected for using data-driven methods. This correction step is only applied on data. The relative residual corrections are based on well-balanced dijet events in which a forward probe jet is calibrated using a tag jet in the well understood barrel region. The last correction step is the absolute residual correction in which reconstructed Z bosons balanced to a jet are used to calibrate the jet energy using the very precisely reconstructed Z boson.

CHAPTER 4

Theory Predictions for the Triple-Differential Dijet Cross Section

Point-like parton-parton scattering processes in high-energy collisions can produce jets with large transverse momenta. Events containing two such jets in the final state, also known as dijet events, allow for rigorous tests of perturbative Quantum Chromodynamics predictions and can subsequently be used to further constrain the proton PDFs and to extract Standard Model parameters like the strong coupling constant α_s.

Next-to-leading-order (NLO) pQCD predictions have been available for dijet and multi-jet observables since many years. These accurately describe shape and normalization of jet cross sections, although they still suffer from larger scale uncertainties limiting the precision of Standard Model parameters extracted from measurements.

For almost ten years, theorists have been working on improving the jet cross section predictions by providing next-to-next-to-leading-order (NNLO) corrections for dijet calculations. When these corrections will become publicly available, they will push the precision of pQCD jet cross section predictions to a new level.

With the finalization of this huge project steadily approaching, we provide a measurement which is specifically designed to benefit from these enhancements. The ultimate aim of this measurement is the improvement of the proton PDFs, especially the gluon PDF, by analyzing dijet events with large transverse momenta.

Common cross section measurements like the inclusive jet cross section are well understood but have the disadvantage that PDF sensitive and insensitive phase space regions are not well separated which limit the constraining power of these measurements.

Therefore, triple-differential cross section observables were studied in this thesis, which promise to yield the most information about the proton structure. As already introduced in Sec. 2.5.1 the rapidities of the dijets are directly related to the fractional proton momenta. So it appears natural, to measure the energy of the dijets as a function of the rapidities of the leading and second jet. However, by explicitly binning the measurement in the leading and second jet as it is suggested in [82], a dependence on the p_{T}-ordering of the jets is introduced. While the ordering according to the transverse momentum of the jets is irrelevant at LO, since both jets are perfectly balanced and have the same transverse momentum, it becomes relevant at NLO. The order of the jets can be changed by a soft emission causing a different ordering of the jets in p_{T}. Simply put, the second jet

can become the leading jet and vice versa. Thus it is not guaranteed that all divergences are canceled. As a consequence the observable becomes infrared unsafe. This can be overcome by filling all events twice into the histograms with interchanged leading and second jet. However, this introduces correlations between phase space regions far off in rapidity and unnecessarily complicates the measurement, especially the unfolding procedure.

Therefore a different definition of the cross section is chosen in this analysis. Instead of using the rapidities and the transverse momenta of each jet, variables which are symmetric between permutations of the leading two jets are used. As these variables are linear combinations of the jet rapidities, no information is lost. Sec. 4.1 presents the cross section definition chosen for the analysis. The subsequent sections present NLO pQCD predictions for the cross section as well as detailed studies of all sources of uncertainties afflicting the cross section calculations.

4.1 Cross Section Definition

The triple-differential dijet cross section is motivated by its relation of kinematic properties of the dijet system to the fractional proton momenta which were introduced in 2.5.1. It is measured as a function of average transverse momentum $p_{\mathrm{T,avg}}$ and is binned in half the absolute rapidity separation, y^*, representing the rapidities of the dijets in the CM frame, and the boost of the dijet system, y_{b}. The triple-differential cross section, which reads as

$$\frac{d\sigma}{dp_{\mathrm{T,avg}}dy^*dy_{\mathrm{b}}}$$

is defined with the observables

$$\begin{aligned} p_{\mathrm{T,avg}} &= \frac{1}{2}(p_{\mathrm{T},1} + p_{\mathrm{T},2}) \\ y_{\mathrm{b}} &= \frac{1}{2}|y_1 + y_2| \\ y^* &= \frac{1}{2}|y_1 - y_2| \end{aligned}$$

where $p_{\mathrm{T},1}$ and $p_{\mathrm{T},2}$ denote the transverse momentum, and y_1 and y_2 the rapidities of the leading and second leading jet. Kinematic cuts of the two leading jets ensure the comparability of the measurement and the calculations:

$$\begin{aligned} p_{\mathrm{T,jet}} &> 50\,\mathrm{GeV} \\ |y_{\mathrm{jet}}| &\leq 3.0 \\ p_{\mathrm{T,avg}} &> 133\,\mathrm{GeV} \end{aligned}$$

The phase space cuts on the transverse momenta and rapidities of the jets are motivated by detector acceptance, trigger efficiencies and furthermore ensure a phase space

with small non-perturbative corrections. The cut on $p_{\mathrm{T,avg}}$ results from the jet triggers employed in the measurement which reach full efficiency only above 133 GeV.

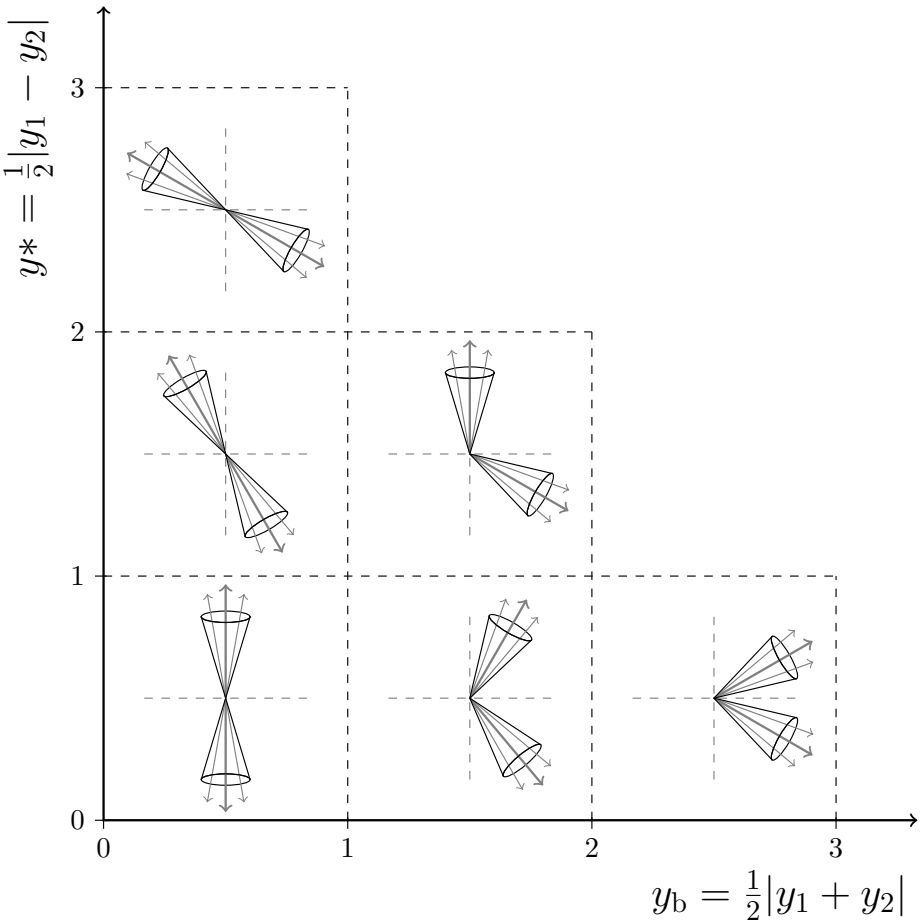

Figure 4.1: An illustration of the dijet topologies in the various y^* and y_b bins. Dijet events are separated in same-side and opposite-side events which allows to draw conclusions about the properties of the initial state partons.

The binning of the cross section in y^* and y_b has the additional advantage that same-side (SS) and opposite-side (OS) dijet events are separated into different bins. Fig. 4.1 depicts the dijet topologies in the various y^* and y_b bins. If the rapidity separation and the boost of the dijet event are small, both jets must have a low rapidity and consequently also small y^* and y_b values. These events are filled in the bottom left bin. If the dijet system is boosted but the two jets have a small separation in rapidity, both jets must be boosted into the forward region to the same side, see bottom right. If instead the dijet separation is large and the boost of the dijet system is small, the jets are boosted to opposite sides. Dijet events with both a large rapidity separation and a large boost of

the dijet system are suppressed in the accessible phase space.

The differentiation in SS and OS dijet events is especially interesting since both event topologies must access different fractional proton momenta while the jets manifest themselves in the same (forward) detector region. Therefore, differences in the predictions can be attributed to the PDFs. This is clearly visible in Fig. 4.6, which shows the PDF uncertainties of the cross section calculations. The measurement bin containing the events with the largest rapidity separation, see bottom right plot, has a significantly larger PDF uncertainty since the high fractional proton momenta are accessed and the PDFs are less precisely known in this region.

4.2 Fixed Order NLO Calculations

The NLO predictions of the triple-differential dijet cross section are calculated using fastNLO [71, 72] which uses interpolation tables filled with the perturbative coefficients of the NLOJet++ program [70], see Sec. 3.4.5. The PDFs are accessed via the LHAPDF library [73, 74] and the α_s-evolution is performed using the routines provided by the PDF sets. Employing fastNLO instead of a direct calculation with NLOJet++ gives the possibility to repeatedly calculate the cross sections with different PDFs and scale choices as it is necessary for the calculation of PDF and scale uncertainties.

4.2.1 Scale Choice

Within perturbative cross section calculations, one has to choose a factorization scale μ_f and a renormalization scale μ_r. The influence of these scales vanishes if the calculation is performed for all orders of the perturbative series. However, since the perturbative series is truncated at NLO, a scale dependence on the result remains. Three possibilities for the scale choice are studied in this thesis. The most natural scale choice is the average transverse momentum of the dijet system which is also used as observable and reflects the energy scale of the measurement.

$$\mu = \mu_\mathrm{r} = \mu_\mathrm{f} = \frac{p_{\mathrm{T},1} + p_{\mathrm{T},2}}{2}$$

While this scale choice yields reasonable results, the k-factors and scale uncertainties indicate problems, which are discussed in detail in Sec. 4.2.2 and Sec. 4.4.1. The second investigated scale choice is based on the findings of a recent analysis by the ATLAS collaboration [83]. They claim that fixed-order calculations which are binned in the rapidity separation y^* become unreliable for high values of y^* if the scale choice only depends on the energy. Based on recommendations of theorists [84], a scale which also depends on the rapidity separation is proposed:

$$\mu = \mu_\mathrm{r} = \mu_\mathrm{f} = p_{\mathrm{T,max}} e^{0.3y^*}$$

Furthermore, a variation of this scale choice is studied in which the scale is not dependent on the transverse momentum of the leading jet but on the average transverse

momentum of the leading two jets as this again resembles the observable in the measurement.

$$\mu = \mu_\mathrm{r} = \mu_\mathrm{f} = p_{\mathrm{T,avg}} e^{0.3y^*}$$

Fig. 4.2 shows the predictions of the NLO calculation using the three discussed scale choices. The cross sections predicted by each calculation are similar with somewhat larger deviations for the scale choice using the maximum transverse momentum instead of the average dijet transverse momentum. The differences between the predictions however are covered by scale uncertainties.

For the final comparisons between data and NLO predictions and for the PDF studies presented later in this thesis, the scale choice $\mu = p_{\mathrm{T,max}} e^{0.3y^*}$ is used, since it exhibits the smallest scale uncertainties.

4.2.2 NLO Correction Factors

To check the influence of higher-order contributions to the perturbative QCD prediction, the differences between LO prediction and NLO prediction are studied, here expressed as the ratio k_NLO.

$$k_\mathrm{NLO} = \frac{\sigma_\mathrm{NLO}}{\sigma_\mathrm{LO}}$$

The size of NLO corrections gives an estimate about the influence of these higher-order corrections. If they are small, the LO result already describes the observable cross section precisely. It is also possible that the k-factors fall below unity, in which case the NLO corrections are negative and the total cross section decreases when adding the correction. Fig. 4.3 shows the k-factors of the NLOJET++ cross section calculations using the discussed scale choices in Sec. 4.2.1. The k-factors are similar in the central region, but the differences increase in regions with larger rapidity separations. Especially the k-factors in the phase space region with a rapidity separation of $2 \leq y^* < 3$ are smaller than unity for the $p_{\mathrm{T,avg}}$ scale choice, while it is larger than unity for the scale choices including the y^* dependence. Interestingly, also the scale uncertainties are much larger in this region.

Apart from the findings for the $p_{\mathrm{T,avg}}$ scale choice, the k-factors are reliable and meet the expectations from previous studies of jet observables.

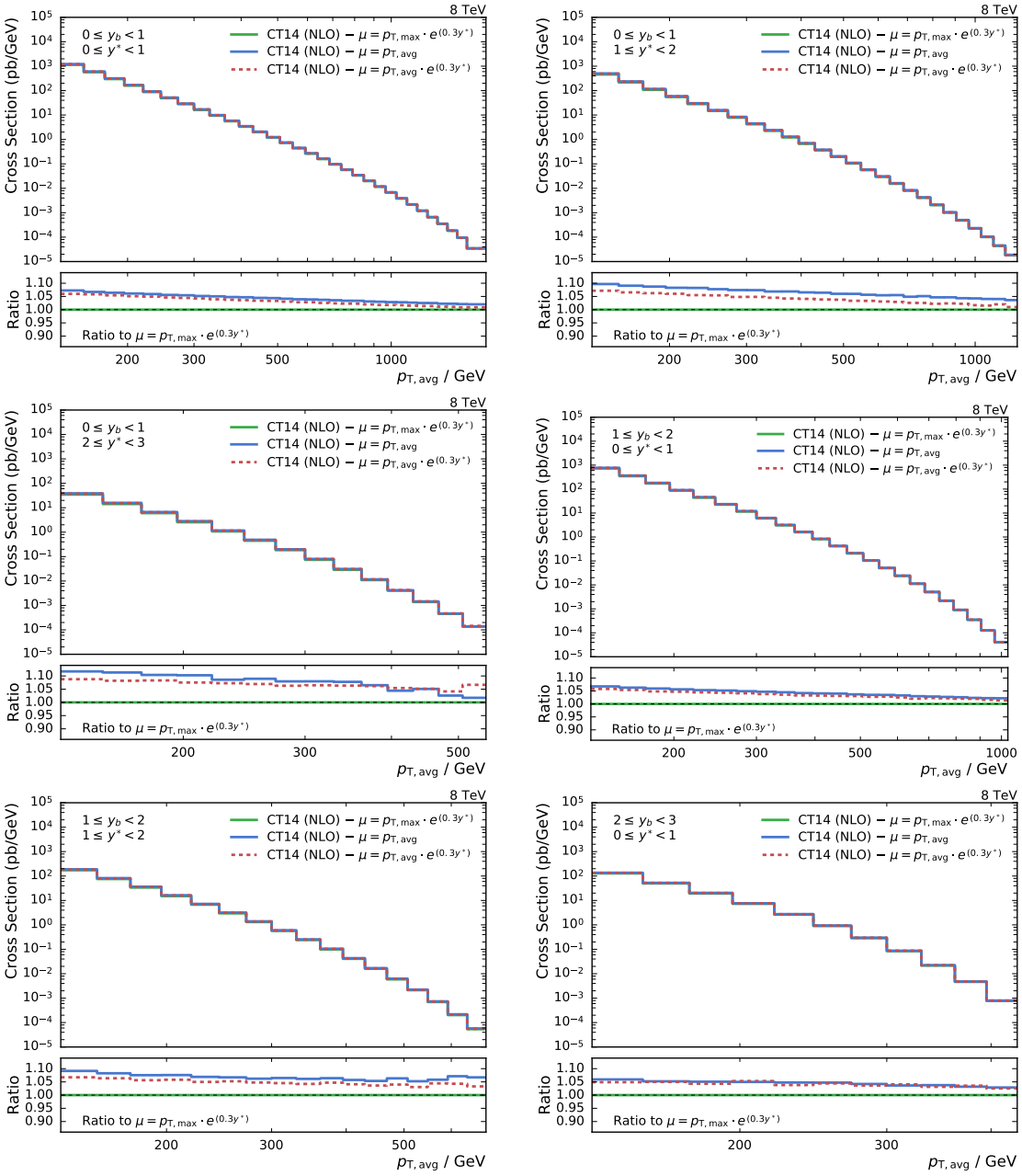

Figure 4.2: NLO predictions of FASTNLO interfaced to NLOJET++ for the triple-differential dijet measurement. The calculations using three different scale choices are shown revealing differences up to 10 %. In almost all cases the differences between the calculations are however covered by the scale uncertainties, see Sec. 4.4.1.

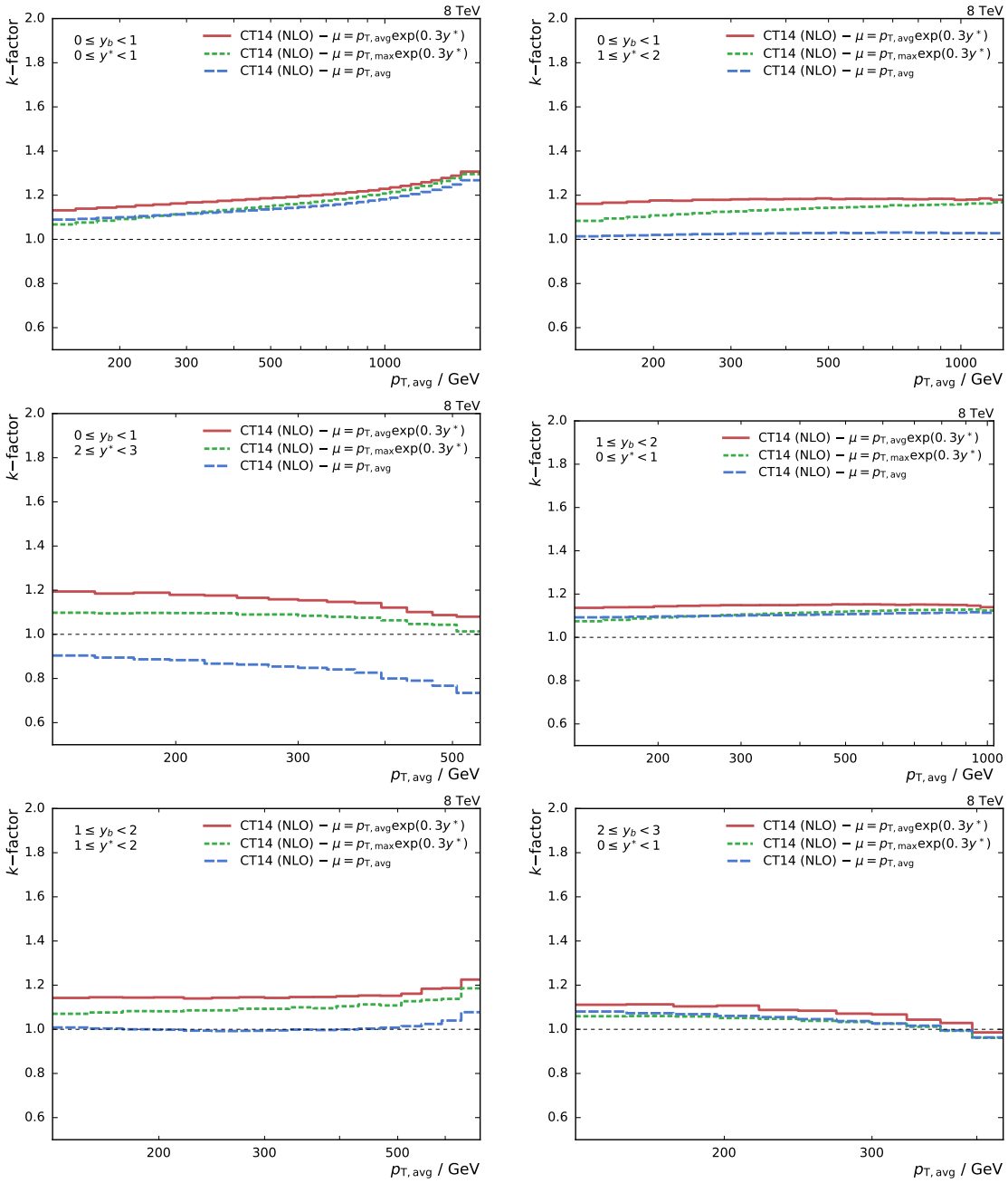

Figure 4.3: The k-factors between the NLO and LO calculation shows the influence of the NLO correction terms. The k-factors for the calculation with $p_{\mathrm{T,avg}}$ as scale choice fall below unity in case of high y^* values indicating that the NLO correction is negative in this phase space region.

4.3 Non-Perturbative Corrections

Perturbative QCD calculations of NLOJET++ give a cross section at NLO parton level. These so-called fixed-order pQCD calculations cannot be directly compared to unfolded data, as they do not include additional soft QCD effects. While a part of these effects is absorbed by the jet algorithm, they still must be estimated and accounted for in comparisons of fixed-order calculations to data.

The influence of these soft effects is estimated using Monte Carlo event generators which are able to simulate those. The non-perturbative (NP) correction c_k^{NP} obtained with a MC event generator k is defined as the ratio between the nominal cross section including multi-parton interactions (MPI) and hadronization effects and a cross section calculation neglecting those, see Eq. 4.3. The superscript in the equation indicates the applied steps in the simulation: the parton shower (PS), the MPI and the hadronization (HAD). The correction is then applied as a bin-by-bin correction factor to the parton-level NLO cross section.

$$c_k^{\mathrm{NP}} = \frac{\sigma^{\mathrm{PS+HAD+MPI}}}{\sigma^{\mathrm{PS}}}$$

The NP corrections have been calculated by employing two LO Monte Carlo event generators using the newest available tunes within CMS. Herwig++ is used with the tune UE-EE-5C [85] and Pythia 8 with the tune CUETP8M1 [86]. Futhermore, the NLO generator Powheg has been used in combination with Pythia 8. Here, the two tunes CUETP8M1 and CUETP8S1 have been studied. For each MC generator, the quotient of a calculation with and without the mentioned effects has been calculated. The ratio is fitted by a power-law function

$$f(p_{\mathrm{T,avg}}) = a \cdot p_{\mathrm{T,avg}}^{b} + c.$$

Since the correction factors obtained from the different MC generators exhibit larger differences, an uncertainty is assigned to the correction factor. The envelope which covers all differences is taken as uncertainty, while the mean of the envelope is used as correction factor.

Fig. 4.4 shows the resulting correction factors and the corresponding uncertainty. The corrections vary between 8 % and 12 % at 133 GeV and get smaller at higher values of $p_{\mathrm{T,avg}}$. While the correction decreases for higher transverse momenta, it does not approach unity, especially in the phase space regions containing boosted jets.

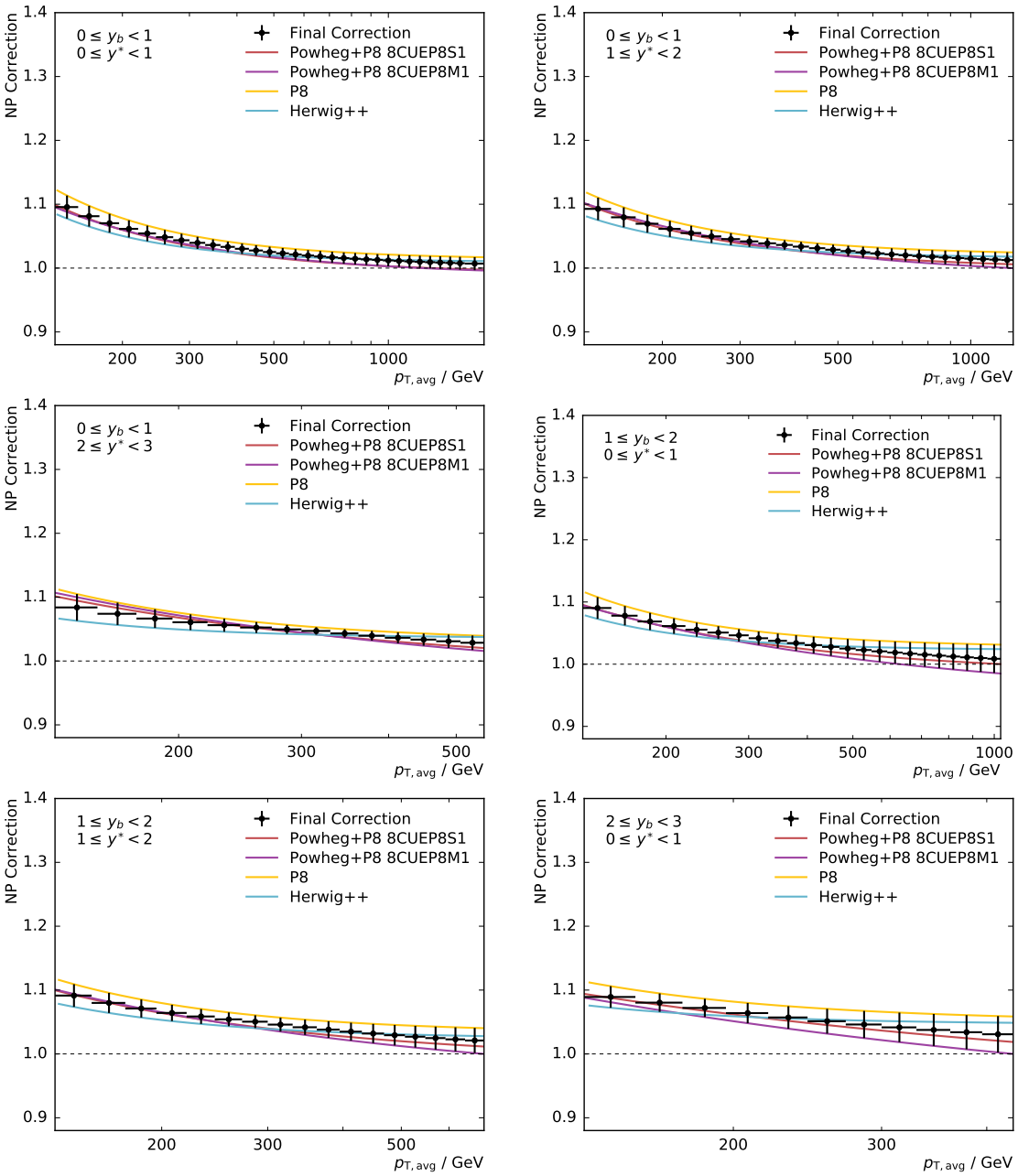

Figure 4.4: The non-perturbative corrections are shown for the six bins in y^* and y_b. They are derived using a MC event generator by calculating the ratio of the cross section with and without hadronization and MPI effects enabled in the simulation. The colored lines show the corrections obtained from different MC generators. The black error bars give the resulting NP correction with uncertainty.

4.4 Theory Uncertainties

Multiple sources of uncertainty limit the precision of the NLO cross section calculation. In this section, the derivation of scale and PDF uncertainties is described. The uncertainties on non-perturbative corrections have been already presented together with the obtained correction factors in Sec. 4.3. A comparison of all studied theoretical uncertainties is shown in Fig. 4.7.

The scale uncertainty is the dominant source of uncertainty in the low-p_{T} region and is of the size of 5 % to 10 % in most of the phase space regions. For larger $p_{\mathrm{T,avg}}$, the PDF uncertainty becomes the dominant source. Its size ranges from 5 % in the low-$p_{\mathrm{T,avg}}$ region up to 50 % for highest $p_{\mathrm{T,avg}}$ and y_{b} values. The uncertainty afflicted to non-perturbative corrections is only sizable for lower p_{T} and is less than 5 % in all bins.

As the cross section calculation is performed using fastNLO interfaced to NLOJet++, multiple independent calculations can be merged to increase the statistical precision of the calculation. The statistical uncertainty is estimated by calculating the uncertainty on the arithmetic mean of the cross section of all fastNLO tables. It is found to be smaller than 0.5 % in all bins, in most of the bins even smaller than 0.1 %. Therefore, the statistical uncertainty is neglected in all further comparisons.

4.4.1 Scale uncertainties

As discussed in Sec. 4.2.1, one has to choose a factorization and renormalization scale for a perturbative cross section calculation and due to the truncated perturbative series, a scale dependence remains. The effects of neglected higher-order contributions are covered by a systematic uncertainty.

There is a common approach in estimating the influence of the scale on the cross section [87]. The cross section calculation is performed using multiple scale choices and the differences compared to the cross section obtained using the central scale choice are translated into a scale uncertainty. The variations are applied as multiplicative factors to the central choice in the following six combinations: $(\mu_{\mathrm{r}}, \mu_{\mathrm{f}}) = (1/2, 1/2), (1/2, 1), (1, 1/2), (1, 2), (2, 1)$ and $(2, 2)$ times the nominal scale. The uncertainty on a quantity X, e. g. the cross section, is the envelope of the maximum deviation in the upwards and downwards direction, while X^0 denotes the value of the quantity with the default scale choice and n is the number of variations.

$$\Delta X^{+} = \max_{i}^{n} \left[X^i - X^0, 0 \right]$$
$$\Delta X^{-} = \max_{i}^{n} \left[X^0 - X^i, 0 \right]$$

Fig. 4.5 shows the relative size of the scale uncertainty for each bin of the measurement and the discussed scale choices. The scale uncertainty is of the size of 5 % to 10 %. However, when using the scale choice $\mu = p_{\mathrm{T,avg}}$, the bin with the largest rapidity separation exhibits a large increase of the uncertainty up to 40 %, which is clearly undesired.

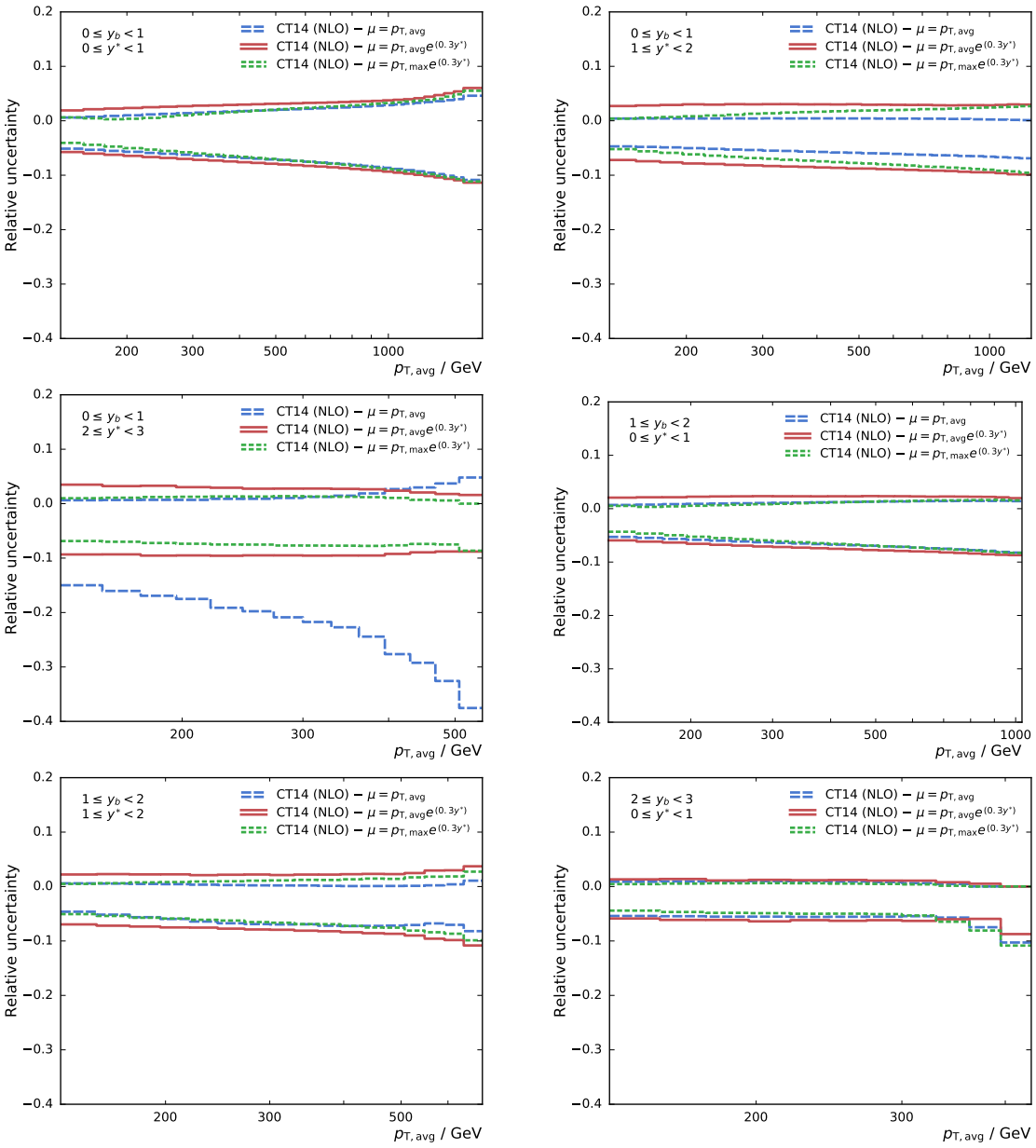

Figure 4.5: The scale uncertainty of the cross section in the six bins of y^* and y_b. The uncertainty is estimated using the common approach of independently varying the renormalization and factorization scale choice in six independent combinations. The uncertainty is shown for the three investigated scale choices, indicated by different colors. In most cases the scale uncertainty is reasonable and of the size of 5 % to 10 %. In the region with the largest values of y^*, the scale uncertainty of the prediction with the scale choice $\mu = p_{T,avg}$ is undesirably large.

4.4.2 PDF uncertainties

The dependence of the cross section calculation on the proton structure is expressed in terms of parton distribution functions, which are derived from fits to data from several experiments. Different sources of uncertainty affect the PDFs. These comprise the choice and the functional form of the parametrization, the chosen theory model and input parameters like the strong coupling constant α_s or the quark masses and, of course also the statistical and systematic uncertainty sources of the data included in the PDF fit.

When determining the PDFs, all these uncertainties are taken into account and are propagated to the PDFs. The groups deriving the PDF sets provide prescriptions how to evaluate these uncertainties. In Sec. 5.9, several comparisons of unfolded data to predictions using the global PDF sets NNPDF 3.0, CT14 and MMHT2014 are shown. In the following, a short summary of the procedure to derive the PDF uncertainties for these PDF sets is given.

The NNPDF PDF set [21] uses a large number of pseudo experiments, in which the PDF fit is performed using data smeared within their uncertainties while taking into account all correlations. These so-called replicas are averaged to give the central result X_{central} and the spread of the replicas determines the uncertainty. The symmetric PDF uncertainties $\Delta X^{\pm}$ of a quantity X, which can be a cross section calculation or even the PDF itself are expressed as

$$\Delta X^{\pm} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left[X_i - X_{\text{central}}\right]^2}$$

where N denotes the number of replicas.

The CT14 [19] and MMHT 2014 [20] PDF sets both employ the eigenvector method to encode their uncertainties. A transformation from the parameter basis to the eigenvector basis is done to yield mutually uncorrelated eigenvectors. By varying the eigenvectors upwards and downwards, a set of eigenvector pairs is generated which can be used to determine the asymmetric uncertainty ΔX^{+} and ΔX^{-} of a quantity X. X_0 denotes the central prediction, X_i^{up} and X_i^{dn} are the predictions using the upwards and downwards variation of the eigenvector PDF set i and N_{EV} is the number of eigenvectors in the PDF set.

$$\Delta X^{+} = \sqrt{\sum_{i}^{N_{\text{EV}}}\left[\max(X_i^{\text{up}} - X_0, X_i^{\text{dn}} - X_0, 0)\right]^2}$$

$$\Delta X^{-} = \sqrt{\sum_{i}^{N_{\text{EV}}}\left[\min(X_i^{\text{up}} - X_0, X_i^{\text{dn}} - X_0, 0)\right]^2}$$

The symmetric uncertainty $\Delta X^{\pm}$ is given by half the difference of the upwards and downwards variation.

$$\Delta X^{\pm} = \sqrt{\sum_{i}^{N_{\text{EV}}}\left[\frac{X_i^{+} - X_i^{-}}{2}\right]^2}$$

The uncertainty assigned to the CT14 PDF set describes a 90 % confidence interval (CI), while the MMHT and NNPDF PDF uncertainties represent a 68 % CI. The CT14 uncertainties are scaled to 68 % CI using $s = \sqrt{2}\,\mathrm{erf}^{-1}(0.9) = 1.645$.

Fig. 4.6 shows the relative PDF uncertainties for the three studied global PDF sets. The PDF uncertainty in the bins with small y^* and small y_b values is comparably small. This is due to the fact that mostly events with opposite side jets contribute, in which the medium-x region of the proton PDFs is accessed which is well known already. The PDF uncertainty of the cross section for high values of y^* and low values of y_b, in which two forward jets are on opposite sides of the detector, is also relatively small. Most interestingly the uncertainty strongly increases for the bin with largest values of y_b. Especially in the high-p_T region, the uncertainty is sizable. To achieve a high boost of the dijet system and a high $p_{\mathrm{T,avg}}$ value, one of the colliding protons must be accessed in the high-x region which is not well determined up to now and is afflicted with large PDF uncertainties. Especially the NNPDF PDF set has a large uncertainty in this region. This is caused by its very flexible parametrization which results in large uncertainties in phase space regions not covered by data.

e vim: tw=80

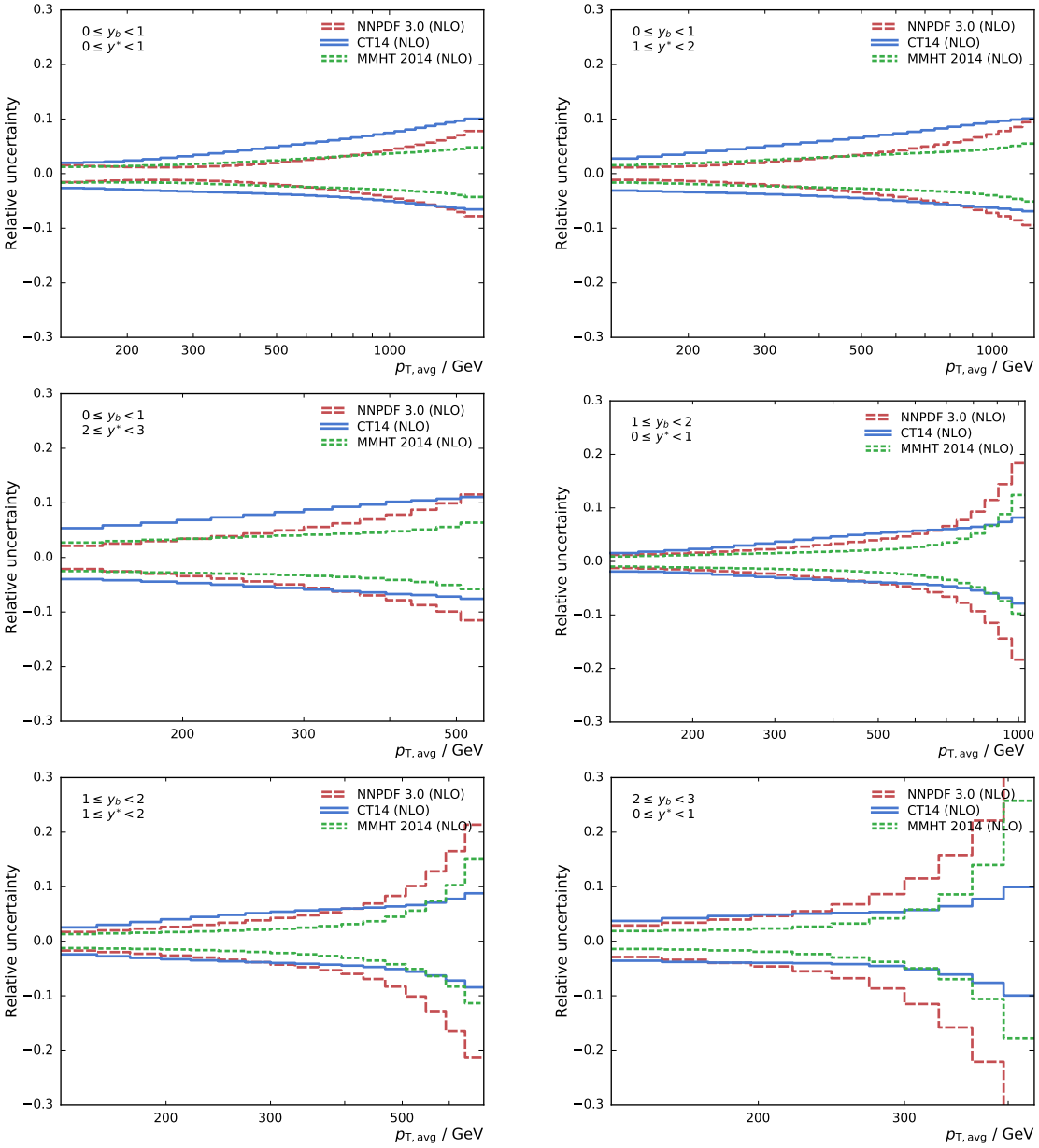

Figure 4.6: The relative PDF uncertainties are shown for the three PDF sets NNDFP 3.0, CT14, and MMHT 2014. The uncertainty represents a 68 % confidence interval. The PDF uncertainty is sizable especially in the boosted region.

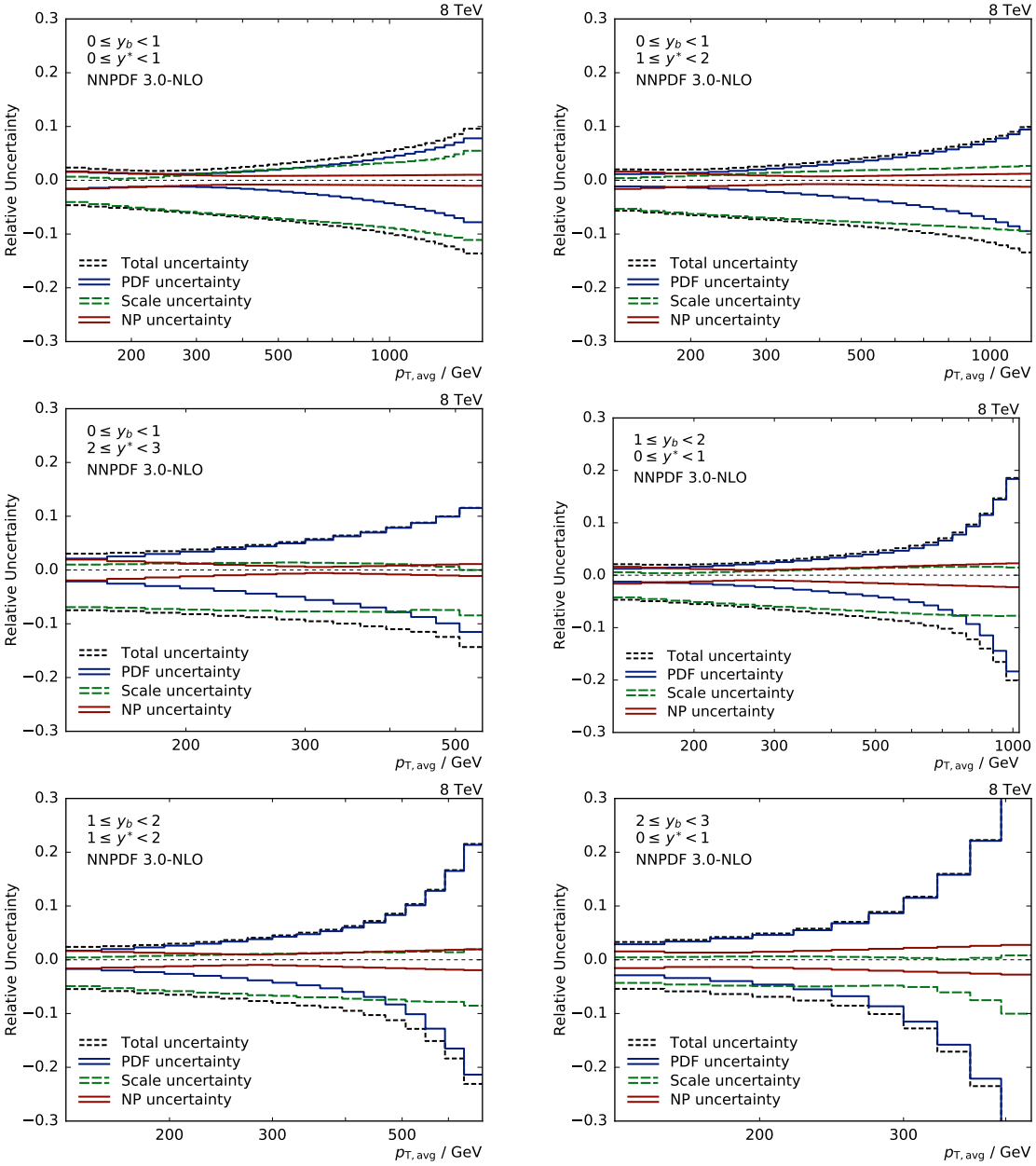

Figure 4.7: Overview of theoretical uncertainties of the NLO predictions. The scale uncertainty is the dominant uncertainty in the low-p_{T} region. At high-p_{T} and especially in the boosted region, the PDFs become the dominant source of uncertainty.

CHAPTER 5

Measurement of the Triple-Differential Dijet Cross Section

The preceding chapter laid the foundation for the triple-differential dijet cross section measurement by presenting the pQCD calculations and demonstrating the observable's sensitivity to the PDFs. Especially the sizable PDF uncertainties in phase space regions involving boosted dijet events indicate that a sufficiently precise measurement could be exploited to better constrain the PDFs.

This chapter is dedicated to the measurement of the cross section from collision data recorded by CMS during the 2012 LHC run at a center-of-mass energy of 8 TeV. The first two sections give the definition of the measured cross section and the analyzed data sets. Section 5.3 describes the event selection in which dijet events are identified. By comparisons with simulated events, the detector response is studied and the jet energy resolution is derived, which is used subsequently to correct the measured cross section for detector effects in an iterative unfolding procedure. A detailed study of all experimental sources of uncertainty in Sec. 5.8 concludes the measurement.

Finally, the comparison of unfolded data to NLO predictions is presented in Sec. 5.9. Furthermore, comparisons to NLO calculations complemented with matched parton showers, as made possible with Powheg and the recently released Herwig 7 MC event generators, are shown.

5.1 Cross Section Definition

The observed dijet event yields are transformed into a triple-differential cross section, which is defined as

$$\frac{d^3\sigma}{dp_{\mathrm{T,avg}}dy^*dy_{\mathrm{b}}} = \frac{1}{\epsilon\mathcal{L}_{\mathrm{int,eff}}}\frac{N}{\Delta p_{\mathrm{T,avg}}\Delta y^*\Delta y_{\mathrm{b}}},$$

where N denotes the number of dijet events, $\mathcal{L}_{\mathrm{int,eff}}$ the effective integrated luminosity and ϵ the product of trigger and event selection efficiencies which are greater than 99% in the measured phase space. The cross sections of background processes like $t\bar{t}$ production are multiple orders of magnitude smaller so that their contributions can be neglected in this analysis. The cross section is normalized by the widths of the $p_{\mathrm{T,avg}}$, y^*, and y_{b} bins.

The size of the $p_{\mathrm{T,avg}}$ bins is chosen based on the observed resolution for $p_{\mathrm{T,avg}}$ at the bin center. For simplicity, the same bin widths as for the inclusive jet cross section measurement are chosen, which has a comparable resolution. The equidistant y^* and y_b bins start from zero and go up to three with a bin size of one.

5.2 Data samples

The measurement is based on data collected by CMS in the 2012 run period of the LHC at a center-of-mass energy of 8 TeV. The accumulated data correspond to a total integrated luminosity of 19.71 fb^{-1}.

The 2012 CMS data taking is subdivided into four periods A, B, C, and D and the data sets are split accordingly into separate samples. Each data set is further split into subsets containing only a fraction of all triggered events, which are grouped with respect to the physics purpose. This analysis relies on a range of jet triggers. All events triggered by prescaled jet triggers were streamed into the data sets `Jet` and `JetMon`, while the events of unprescaled jet triggers were stored in the data set `JetHT`. As this assignment changed over the 2012 run periods, it was taken care to only select the correct subset of events. Table 5.1 presents the CMS collision data sets analyzed in this thesis as well as the integrated luminosity collected in each run period.

Table 5.1: The 2012 collision data of CMS comprise four data sets collected in the run periods A, B, C and D. The total integrated luminosity of all data collected in 2012 sums up to 19.71 fb^{-1}.

Run	Run range	Data set	Luminosity fb^{-1}
A	190456–193621	/Jet/Run2012A-22Jan2013-v1/AOD	0.88
B	193834–196531	/Jet[Mon,HT]/Run2012B-22Jan2013-v1/AOD	4.41
C	197770–203755	/Jet[Mon,HT]/Run2012C-22Jan2013-v1/AOD	7.06
D	203773–209465	/Jet[Mon,HT]/Run2012D-22Jan2013-v1/AOD	7.37

5.2.1 Monte Carlo Event Samples

To compare the measured data with simulated events, two Monte Carlo event generators were used, which are described in Sec. 3.4.5. Each MC generator is utilizing an optimized tune to simulate the underlying event. Additionally, the Monte Carlo event samples contain not only the hard scattering event, but also an admixture simulating the pileup collisions observed in real collisions.

The Madgraph Monte Carlo sample is generated with a multijet QCD process. The LO matrix elements do not only contain the $2 \to 2$ matrix elements, but also tree-level multijet matrix elements. These better describe multi-jet events, but have to be matched to the subsequent parton shower. The underlying event is modeled using the tune Z2*. The parton shower and hadronization is carried out using the Pythia 6 event generator interfaced to Madgraph by the LHE event record [88]. This simulated event sample is

the main sample used for data comparisons, not because of the multileg-improved matrix elements, which are not as important for the dijet analysis in this thesis, but because of the huge number of events generated and simulated.

Furthermore, a data sample generated with Pythia 8 is used. While including only the $2 \to 2$ LO matrix elements, the event simulation and the underlying event tune are improved in this newer version of the Pythia Monte Carlo generator.

To be able to perform comparisons between data and simulated events on reconstructed level, events from both MC event generators were propagated through the complete simulation of the CMS detector. This allows for further studies like the measurement of the jet transverse momentum resolution, which are needed to finally correct the measured cross sections for detector effects.

Predictions of jet cross sections, especially those involving the jet transverse momenta, are notoriously difficult to be calculated because of the steeply falling spectrum. Consequently, it is not possible to generate enough events to populate all the phase space up to highest p_{T} with a sufficiently large number of events. However, there are two approaches to overcome this issue. The first method involves the reweighting of the generation process to generate more events at high transverse momentum. Unfortunately, this spoils the absolute normalization of the cross section as well as it involves the consideration of huge event weights. The second method is based on splitting the phase space into multiple regions, in which events are generated separately. The different phase space regions are then stitched together in the data analysis while taking into account the cross sections in the different phase space regions.

The Madgraph sample is split into four regions according to the scalar sum of the transverse momenta, H_{T}, while the Pythia 8 sample is split into twelve data sets according to the transverse momentum of the leading jet. Table A.1 shows all Monte Carlo data sets as well as their cross section and the generated number of events.

5.3 Event selection

The selection of events is based on several quality criteria which are either recommended by CMS or developed specifically for this analysis in order to yield a dijet event sample with high purity and high selection efficiency. Furthermore, phase space cuts on the jets ensure the applicability and comparability of NLO theory calculations.

5.3.1 Certified Data Selection

The first step in the event processing chain is to only pick data from runs and lumi-sections[1]which fulfill certain criteria. These criteria include proper performance of all detector subsystems as well as the passing of data quality monitoring (DQM) steps during the validation process. The good sections within a run are announced using a data file. The applied certification file[2] in this analysis is based on the final event reconstruction

[1]CMS stores data split in different subsets according to a fixed time range in which the instantaneous luminosity is assumed to be constant.

[2]`Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON`

of the 2012 CMS data sets.

5.3.2 Trigger Selection

To measure and reconstruct the $p_{\mathrm{T,avg}}$ spectrum of dijet production, a set of single jet triggers has been used. Single jet triggers consist of one L1 trigger seed and multiple HLT filters. The L1 trigger has a lower threshold to ensure full efficiency versus p_{T} of the HLT trigger. Since the p_{T} spectrum is steeply falling and the rates for low-p_{T} jets are very high, it is not feasible to use a single unprescaled trigger for the selection of all interesting events. Instead, a set of five prescaled low-p_{T} trigger paths, each with different prescale value, is used to collect sufficient data in the lower part of the p_{T} spectrum. Additionally, one unprescaled trigger is used in the high p_{T} region, in which the rate is sufficiently small to collect and store all events. A prescale value of n means, that only every n^{th} event accepted by a trigger, is kept. Of course, the applied prescale values are taken into account when the original spectrum is reconstructed.

Table 5.2: List of all single jet trigger paths used in the analysis. Each trigger is employed in a mutually exclusive phase space region in $p_{\mathrm{T,avg}}$, in which the trigger exhibits an efficiency larger than 99%. The column $p_{\mathrm{T,avg,99\%}}$ indicates the value at which each trigger reaches that value. The last column reports the effective luminosity seen by each trigger. This number, divided by the total integrated luminosity of 19.71 fb^{-1}, gives the effective prescale applied on a trigger over the whole run period.

Trigger path	**L1 threshold** GeV	**HLT threshold** GeV	$\boldsymbol{p_{\mathrm{T,avg,99\%}}}$ GeV	**Eff. Lumi** fb^{-1}
HLT_PFJet40	16	40	—	0.8×10^{-4}
HLT_PFJet80	36	80	123	0.21×10^{-2}
HLT_PFJet140	68	140	192	0.56×10^{-1}
HLT_PFJet200	92	200	263	0.26
HLT_PFJet260	128	260	353	1.06
HLT_PFJet320	128	320	412	19.71

Table 5.2 shows all single jet triggers used in the data analysis. Each trigger is used in mutually exclusive regions of $p_{\mathrm{T,avg}}$, in which the trigger is fully efficient.

The jet reconstruction algorithms and the jet energy corrections applied on HLT level slightly differ from the ones used for the final event reconstruction. Furthermore, the efficiency of each trigger is not calculated as a function of jet p_{T}, which was used in the trigger decision, but versus $p_{\mathrm{T,avg}}$ of the two leading jets as used in this analysis. Therefore, the triggers exhibit a turn-on behavior, as can been seen in Fig. 5.1. Consequently, it is necessary to determine the threshold above which a trigger becomes fully efficient. It is defined as the value at which the efficiency exceeds 99%.

Basically, it is possible to calculate the efficiency of a given trigger by dividing the number of passing events through the number of events that pass the next-lower trigger in p_{T}, because by definition the looser trigger is efficient, as soon as the higher trigger becomes efficient. This is aggravated through the different prescales applied to each

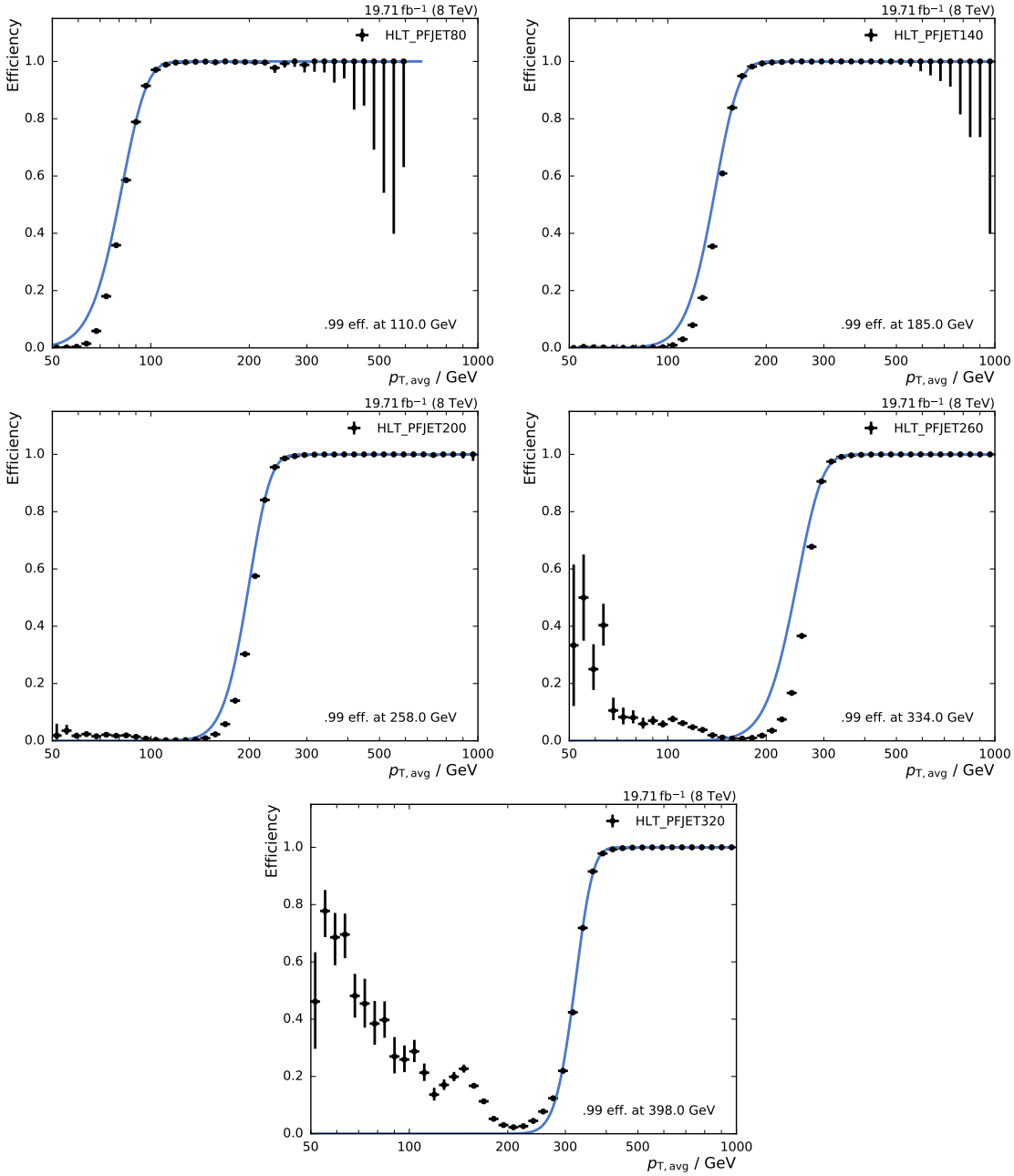

Figure 5.1: Trigger turn-on curves for the single jet trigger paths used in the analysis. To determine the 99% efficiency threshold, the trigger turn-on curves are fitted using a sigmoid function taking into account the uncertainties using Clopper-Pearson confidence intervals.

trigger path. While it is possible to normalize the yield by the effective luminosity seen by each trigger, this method is affected by larger statistical fluctuations as the number of events differs strongly between the two trigger paths.

Therefore, a more challenging but superior method is employed: When the L1 and HLT triggers are processed, the jet four-vectors, on which the trigger decision is based, are stored. Thus, it is possible to recalculate the trigger decision by comparing the transverse momentum of the L1 trigger object with the L1 threshold and the p_{T} of the HLT trigger object with the HLT threshold [52].

Similarly, the trigger decision of the next higher trigger can be emulated starting from a lower trigger path. A set of events $S_1 = \{E_i | T_A(E_i)\}$ which was accepted by the lower trigger path T_A is used to determine the subset $S_2 = \{E_i | T_A(E_i) \wedge T_B(E_i)\}$ which also passes the next higher trigger T_B, see Eq. 5.1. The quotient of both event sets is used to determine the turn-on curve as shown for each trigger path in Fig. 5.1. The uncertainty on the efficiency is indicated by error bars which represent Clopper-Pearson confidence intervals.

$$f_{\mathrm{eff}}(x) = \frac{N(\{E_i | T_A(E_i) \wedge T_B(E_i), x)\}}{N(\{E_i | T_A(E_i)\}, x)} \tag{5.1}$$

To determine the point, at which the trigger efficiency is larger than 99%, the turn-on distribution is fitted using a sigmoid function that describes the turn-on behavior of the trigger paths close to the efficiency threshold.

$$f_{\mathrm{fit}}(x) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right) \tag{5.2}$$

The thresholds which were finally used in the data analysis deviate slightly from the ones shown in Fig. 5.1 as the trigger thresholds were measured separately in each y^* and y_{b} bin. The most conservative, thus the highest threshold, is finally chosen. They are reported in Table 5.2.

It is important to mention the HLT triggers which specifically trigger on the average transverse momentum of the two leading jets and have been implemented for dijet calibration purposes. At first, they appear to be the obvious choice for this analysis. However, studies of these triggers revealed, that the performance is at best comparable to the single jet triggers. Larger prescales applied to the $p_{\mathrm{T,avg}}$ trigger paths result in larger statistical uncertainties. Therefore, the single jet trigger paths were used in this analysis.

5.3.3 Primary Vertex Selection

The cuts on the primary vertices further reject beam background and off-center bunch crossings. Each event has to contain at least one primary vertex (PV) which is well reconstructed within a distance of $|z_{\mathrm{PV}}| < 24\,\mathrm{cm}$ to the nominal interaction point of the detector. Furthermore, the radial distance ρ_{PV} needs to be smaller than 2 cm. To ensure a high quality of the vertex reconstruction, the number of degrees of freedom in the

vertex fit, $n_{\mathrm{dof,PV}}$, needs to be at least four. Thus, at least four tracks must be present in order to perform a valid vertex fit.

5.3.4 Missing Transverse Energy Cut

If all particles could be identified and perfectly measured, the transverse momentum of all particles would sum up to zero. The imbalance in the transverse momentum of all visible particles, which can be measured in the detector, is called the missing transverse momentum (MET). Neutrinos, for example, leave the detector undetected and contribute to the MET. MET is an important ingredient in many measurements involving W bosons, top quarks or searches for physics beyond the Standard Model which involve undetectable particles.

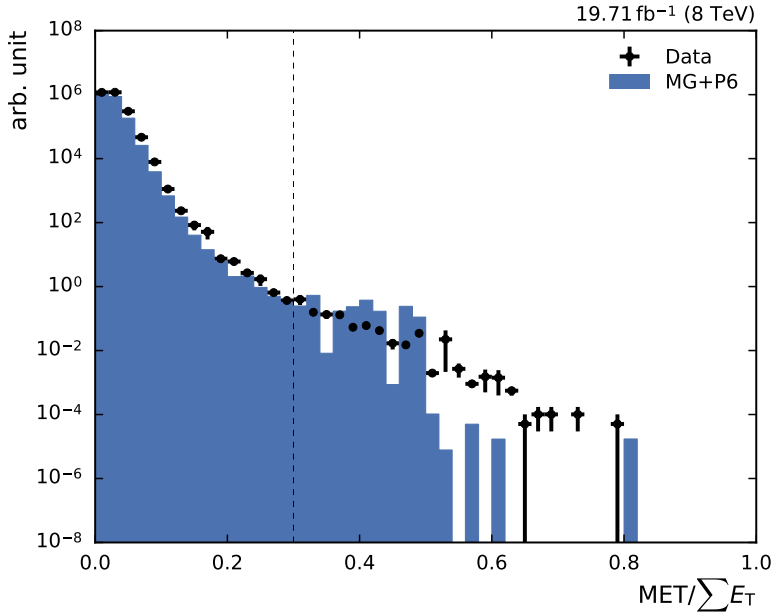

Figure 5.2: Missing transverse energy fraction of the total transverse energy per event in data and simulated events. To remove background and noise, events with a fraction exceeding a certain threshold, here indicated by the dashed line, are rejected.

However, a large fraction of MET in an event is not always caused by interesting physics processes. Very often, the reason can be found in detector noise, cosmic rays or beam-halo particles. Therefore, a sequence of algorithms developed by the MET working group at CMS [89] is employed, which identifies and rejects these events. Moreover, a cut removing events in which the missing transverse energy fraction $\not{E}_{\mathrm{T}}$ constitutes a large fraction of the total transverse energy $\sum_i E_{\mathrm{T},i}$ is applied, see Fig. 5.2,

$$\frac{\not{E}_{\mathrm{T}}}{\sum_i E_{\mathrm{T},i}} < 0.3. \tag{5.3}$$

5.3.5 Jet Identification

The jet identification criteria (jet ID) reject noise and noise-enhanced jets while all real jets are kept. The jet ID is not applied per event, but each jet is accepted or removed from the list of valid jets. The algorithm works on reconstructed jets using information of the clustered particle candidates. Following the official recommendations of the JetMET group [90], the so-called loose jet ID is used. All jets passing the jet ID are then further processed in the analysis chain.

The properties of the reconstructed jets and their respective cuts are listed in Table 5.3. The cut on the fraction of neutral hadrons and photons removes HCAL noise and ECAL noise, respectively. Muons that are falsely identified and clustered as jets are removed by the muon fraction criterion. Based on information of the tracker, additional selection cuts are enforced in the region $|\eta| < 2.4$. Jets clustered from misidentified electrons are removed by the charged electromagnetic fraction cut. Furthermore, the fraction of charged hadrons in the jet must be larger than zero. This cut is important since the CHS algorithm removes charged particles from pileup vertices. Consequently, jets without any charged hadrons are very likely to be pileup jets. The Figs. 5.3 and 5.4 show the distributions of the jet constituents observed in data and simulated events.

While studying the loose and tight jet criteria, it was found that the tight jet ID removes a non-negligible fraction of jets in the forward region ($|y| > 2.4$) of the phase space considered in this analysis. Therefore, the loose jet ID is favored and applied in this thesis which is also the official recommendation of the JetMET group.

Table 5.3: The jet ID removes noise and fake jets based on the properties of the reconstructed jets and the clustered particle candidates. All selection cuts which are recommended by the JetMET group are applied [90]. The loose jet ID is used in this analysis as the tight ID removes a non-negligible fraction of signal events, particularly in the forward region without tracker coverage.

	Property	**Loose ID**	**Tight ID**
Whole η region			
	neutral hadron fraction	< 0.99	< 0.90
	neutral EM fraction	< 0.99	< 0.90
	number of constituents	> 1	> 1
	muon fraction	< 0.80	< 0.80
only $\lvert\eta\rvert < 2.4$			
	charged hadron fraction	> 0	> 0
	charged multiplicity	> 0	> 0
	charged EM fraction	< 0.99	< 0.90

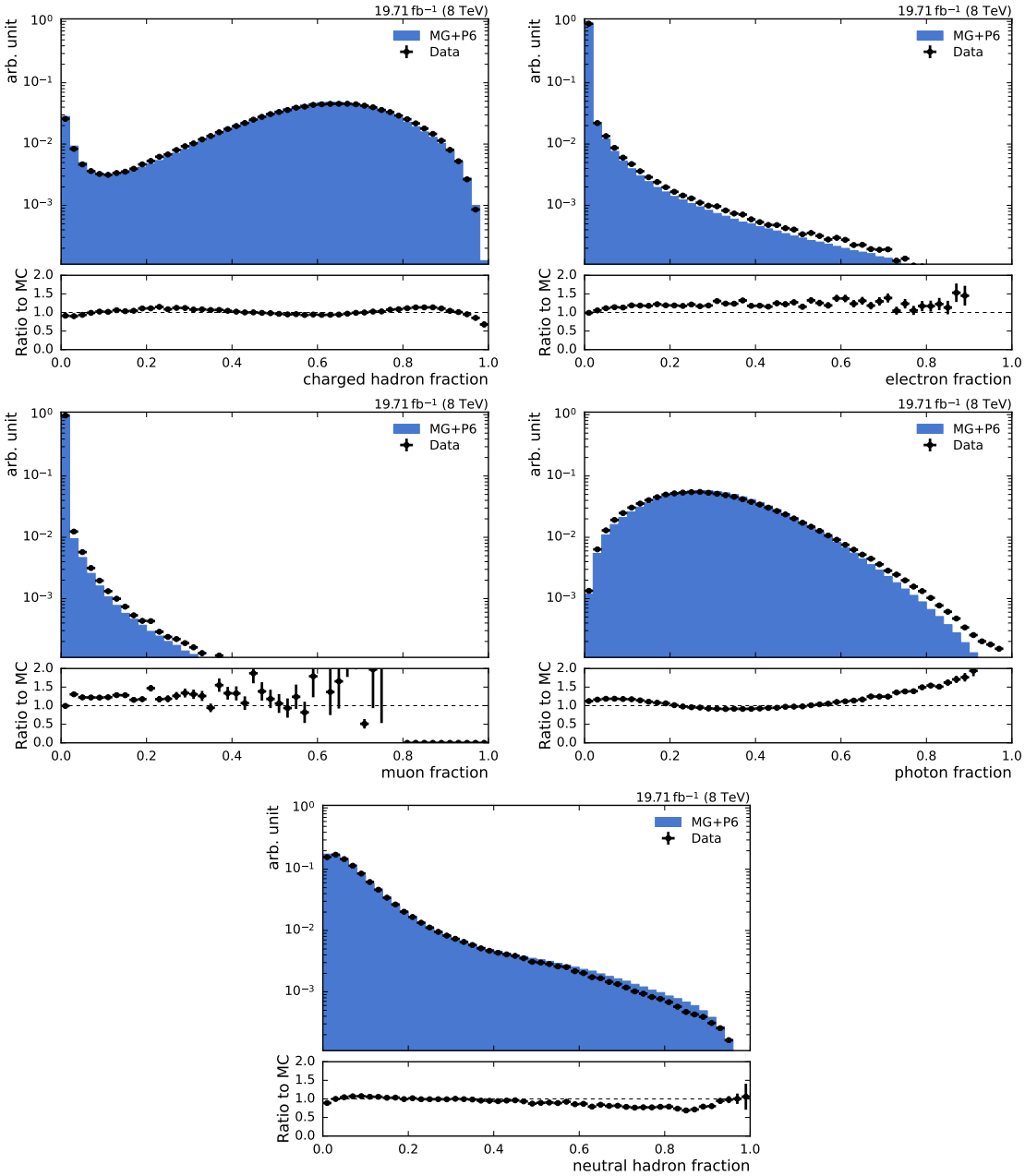

Figure 5.3: The fractions of jet constituents as observed in data and simulated events for different types of PF candidates. Data and simulation are normalized to the same number of events. The distributions are shown after the application of the jet ID.

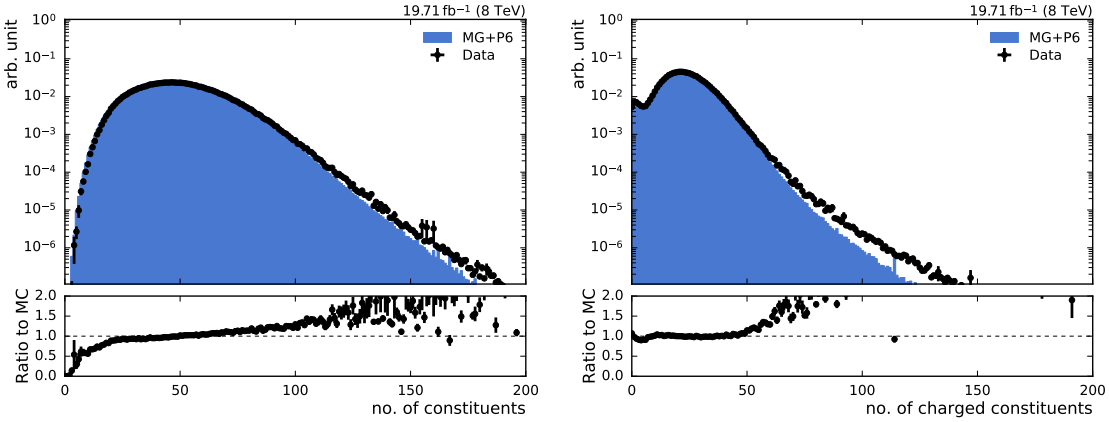

Figure 5.4: Number of PF candidates clustered into a jet in data and simulated events. Data and simulation are normalized to the same number of events.

5.3.6 Jet ID Efficiency

The applied jet ID ensures a high purity of real jet events. The efficiency of the jet ID in the investigated phase space is studied using a tag-and-probe technique. Dijet events, in which the leading two jets are well balanced in ϕ, are selected using

$$|\Delta\phi - \pi| < 0.3,$$

where $\Delta\phi$ is the azimuthal separation between the two leading jets. One of the dijets is chosen as tag jet and has to fulfill the loose jet ID. For the other jet it is examined, whether it also passes the jet ID. The quotient of events in which the probe jet also passes the criteria versus the total number of dijet events yields the efficiency. Fig. 5.5 shows the efficiency as a function of $p_{\mathrm{T,avg}}$ of the dijet system for all y^* and y_b bins. As advertized by the JetMET group, the efficiency is larger than 99 % in all phase space regions.

5.3.7 Jet Energy Corrections and Selection

The measurement presented in this thesis is based on jets clustered from PF candidates using the anti-k_T jet algorithm with a size parameter of 0.7. The following phase space cuts remove jets instead of whole events due to their transverse momentum and rapidity. Consequently, all jet energy corrections recommended by CMS are applied prior to this selection in order to have the correct energy scale of the jets. The factorized correction approach employed by CMS is discussed in Sec. 3.5.4 and comprises different correction levels for jets in data[3] and for jets in simulated events[4].

The accessible phase space in theoretical calculations and in the measurement is synchronized by selecting jets only from a restricted part of the complete phase space, in

[3]The JEC version applied on data is internally referred to as `Winter14_V8`
[4]The latest JEC for run-independent Monte Carlo Samples are called `START53_V27`

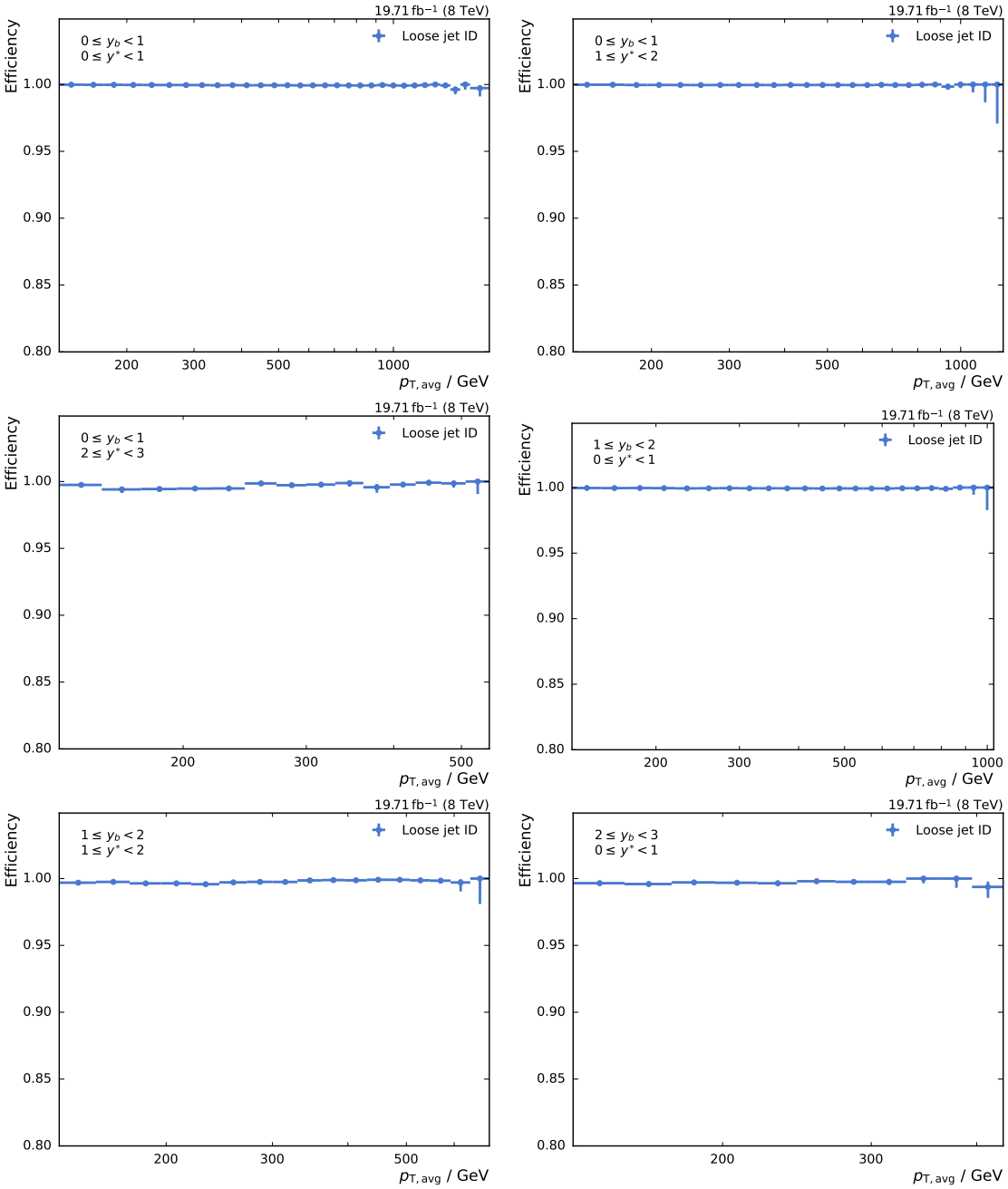

Figure 5.5: The jet ID efficiency as a function of $p_{\mathrm{T,avg}}$ for all y^* and y_{b} bins. It is studied using a tag-and-probe approach on dijet event topologies. The efficiency is shown as a function of $p_{\mathrm{T,avg}}$ for all bins. It always exceeds 99 %.

which the detector acceptance is high and the applicability of NLO pQCD calculations is guaranteed:

$$p_{\mathrm{T,jet}} > 50\,\mathrm{GeV}$$
$$|y_{\mathrm{jet}}| \leq 3.0$$

Events in which the leading or second jet fail the jet selection are discarded in order to only keep events in which both leading jets stem from the hard scattering. Furthermore, an additional cut on the average transverse momentum of the dijets is applied:

$$p_{\mathrm{T,avg}} > 133\,\mathrm{GeV}$$

This cut is necessary, as the first employed single jet trigger becomes efficient at this point. Often, it is recommended to have asymmetric cuts on the transverse momentum of the two leading jets to avoid an infrared sensitive region. Because of the much higher cut on the average transverse momentum, this is not an issue here. Other advantages of the high cut on $p_{\mathrm{T,avg}}$ are the avoidance of a turn-on region which would lead to complications in the applied unfolding procedure as well as the restriction to a phase space region in which non-perturbative contributions are small.

5.3.8 Angular Jet Corrections

The jet energy corrections relate the reconstructed jet energy to the particle-level jet energy, but they do not include any correction for angular reconstruction biases of the jets. However, especially in the transitional regions of the detector, i. e. when the tracker coverage ends, a systematic reconstruction bias is observed, see Fig. 5.6. Jets are reconstructed with systematically larger absolute pseudorapidity, meaning they are shifted towards the forward region. While the absolute shift is rather small, it causes a relevant systematic effect whenever the rapidity separation of two jets is of interest. There are two effects however, that limit the impact of the misreconstruction: First, it is only pronounced for jets with low transverse momentum (as can be seen in Fig 5.7) and second, the large bin size in y^* and y_{b} used in this analysis reduces the impact.

Nonetheless, the systematic bias is accounted for by applying a correction on each jet based on the average difference between the pseudorapidity of particle-level jets and reconstructed jets as a function of reconstructed jet p_{T} and pseudorapidity. The Figs. 5.6 and 5.7 also show the distribution after applying the correction, in which the majority of the η-dependent effects is removed.

To estimate the effect on the resulting cross section, a comparison between the yielded cross section with and without applying the angular correction is presented in Fig. 5.8. In the bins containing boosted dijets, the correction causes changes of about 2 % of the cross section. Since these events move to bins containing more central jets in which the cross section is much higher, no change is noticeable there.

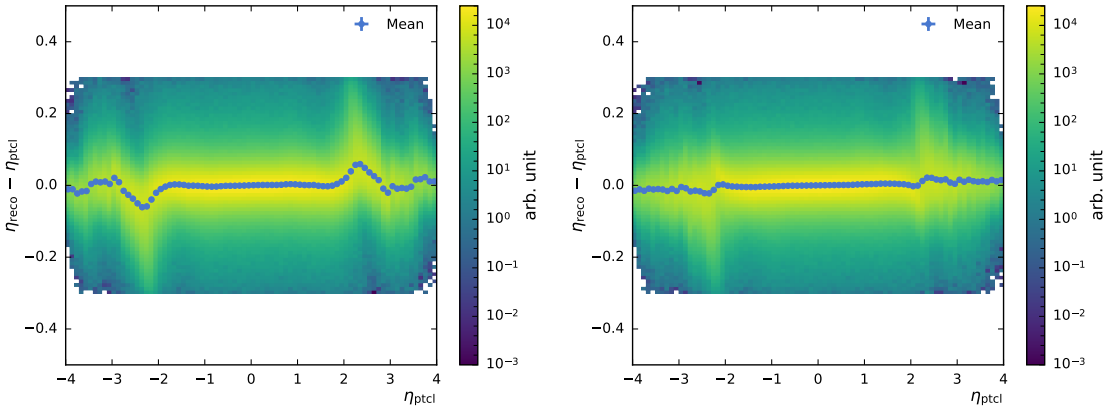

Figure 5.6: The difference between the pseudorapidity of the reconstructed jets and the generated jets is shown over the pseudorapidity of the generated jet. The distribution is shown before (left) and after (right) applying the pseudorapidity correction. The blue points indicate the mean of the distribution in each η_{gen} bin.

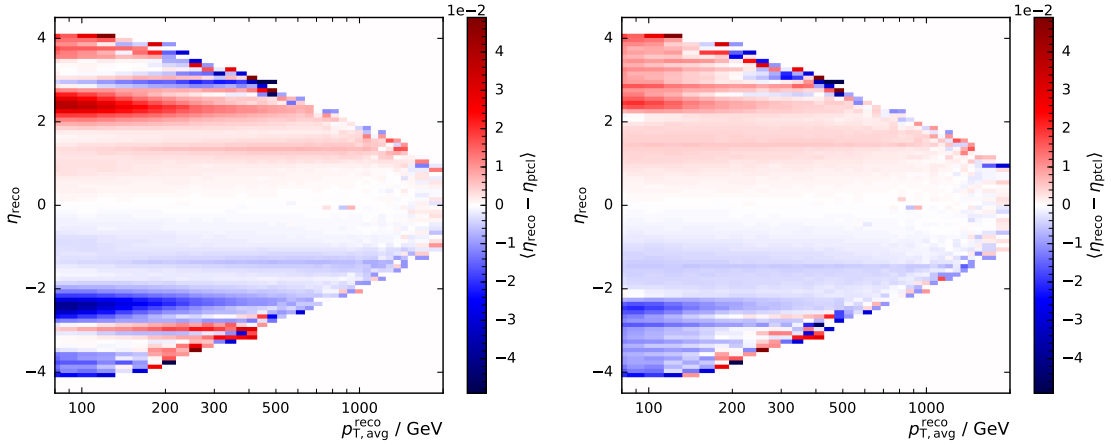

Figure 5.7: The pseudorapidity of the generated jets is shown as a function of the transverse momentum of the jets. The color indicates the differences of the pseudorapidity of the generated and reconstructed jets. The distribution is shown before (left) and after (right) the angular correction is applied.

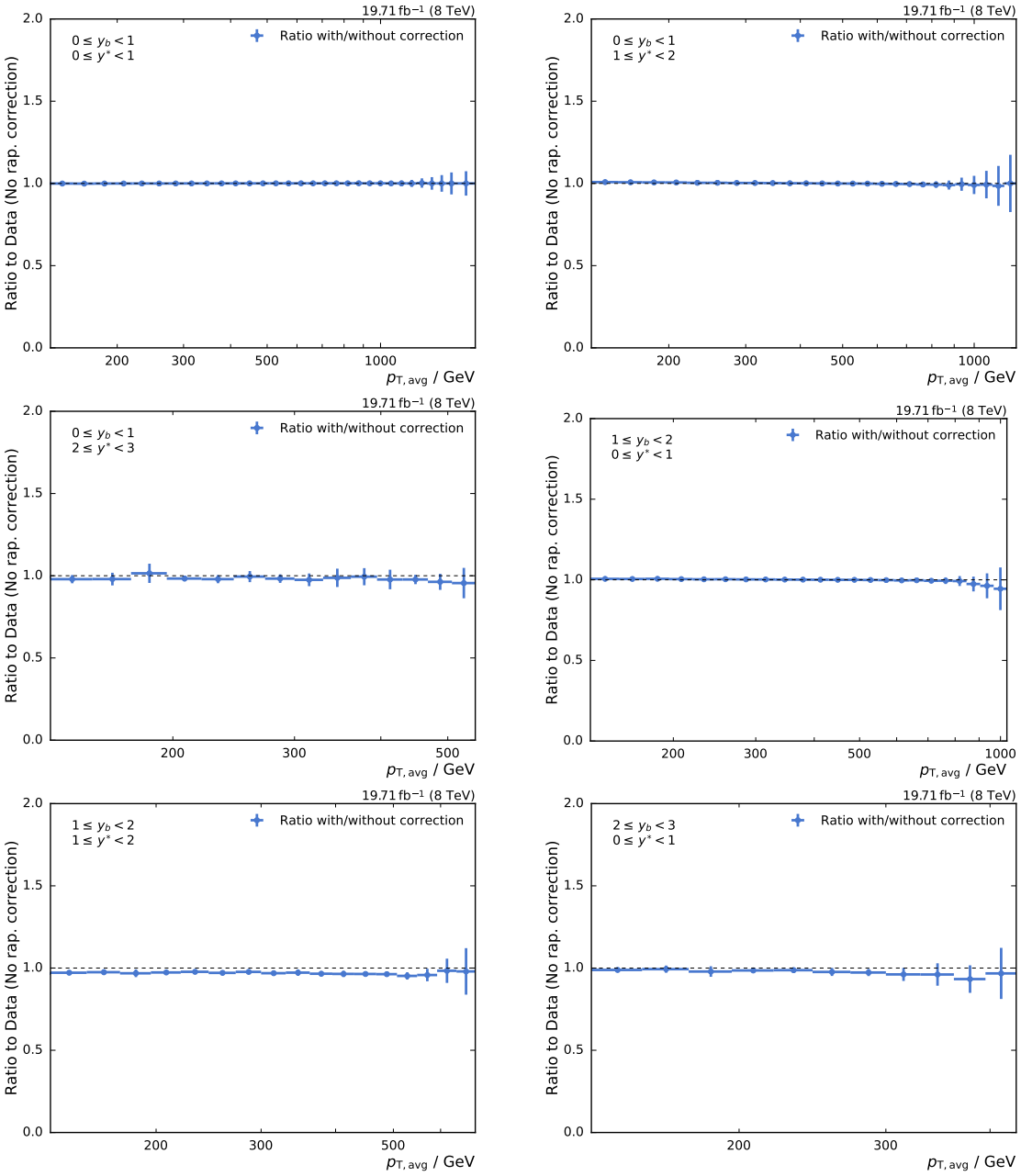

Figure 5.8: The effect of the angular correction on the cross section is demonstrated by calculating the ratio of the cross section after applying the correction to the one without correction. The cross section decreases in bins involving forward jets and is more pronounced for higher values of y^*.

5.3.9 Stability versus Run Periods

The experimental conditions for data-taking change slightly over the various run periods due to changes of the detector calibration or different trigger prescales. Nonetheless, the measured cross sections must not depend on these effects. This is studied by analyzing the different run periods separately. The result for each run period is shown as ratio to the cross sections obtained in run D in Fig. 5.9.

There are differences visible, most notably due to statistical fluctuations in the high-p_{T} region. Furthermore, a slight slope of the cross section obtained in run B is observed. However, the results are in agreement within uncertainties.

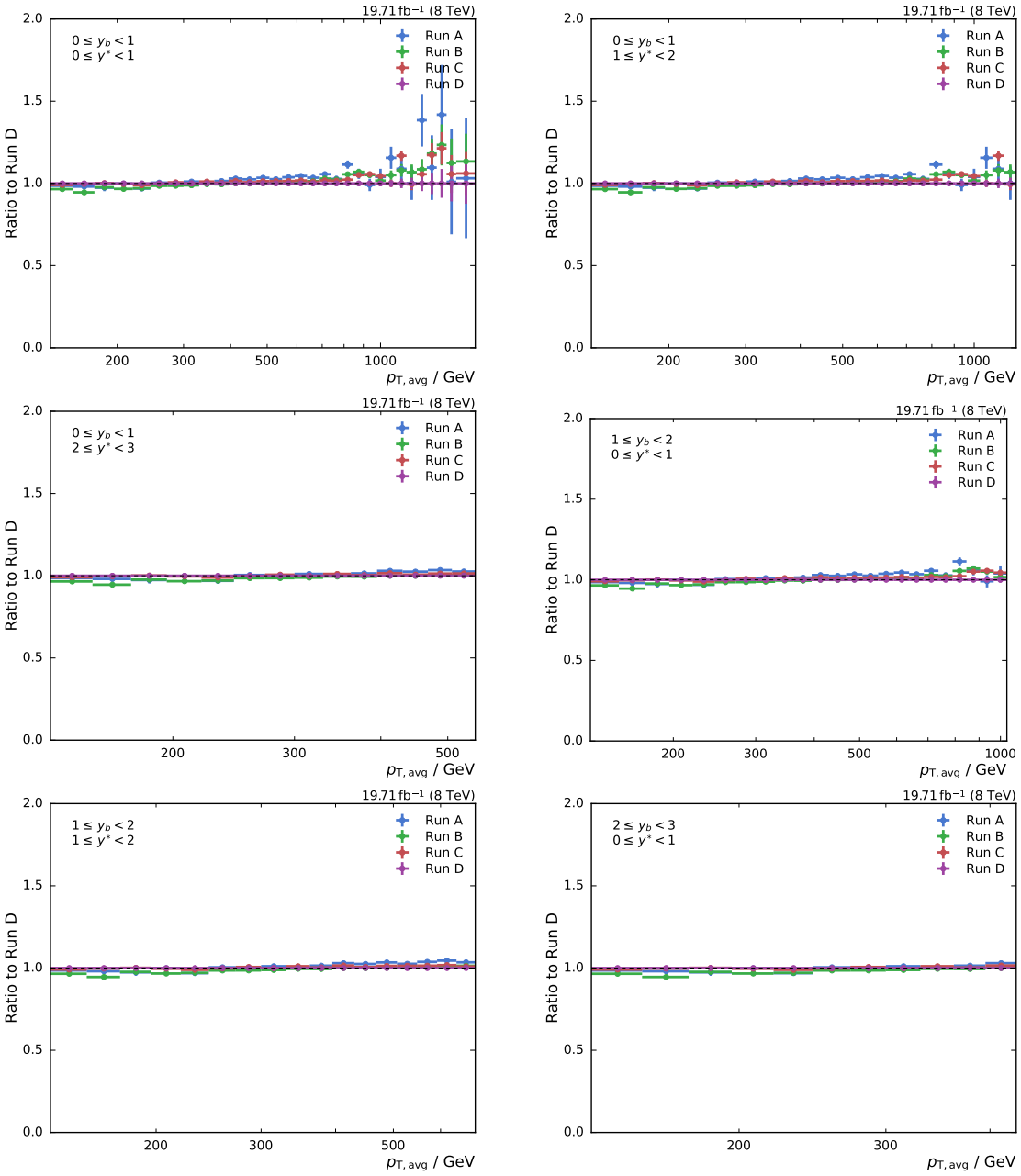

Figure 5.9: Ratio of the measured cross section in each run period to the cross section obtained with data from run D. There are small differences, mostly due to statistical fluctuations. A slight slope is observed for the result of run B, but overall the separately obtained cross sections are in agreement within uncertainties.

5.4 Comparison with Simulated Events

5.4.1 Pileup Reweighting

The official Monte Carlo samples are enriched with an admixture of pileup collisions to mimic the pileup distribution expected in data. Ideally, the estimated pileup distribution in data $N_{\text{data}}(N_{\text{PU,est.}})$ would match with the simulated distribution $N_{\text{MC}}(N_{\text{PU,truth}})$. Since the admixture is only a rough estimate of the pileup distribution expected in the forthcoming data taking, a perfect matching cannot be achieved. To still get comparable pileup distributions in data and simulated events, the simulated events are reweighted with w_{PU} to match the distribution in data:

$$w_{\text{PU}}(N_{\text{PU,truth}}) = \frac{N_{\text{data}}(N_{\text{PU,est.}})/\sum N_{\text{data}}}{N_{\text{MC}}(N_{\text{PU,truth}})/\sum N_{\text{MC}}}$$

Fig. 5.10 shows the number of reconstructed primary vertices before and after reweighting. The significant mismatch of the pileup distributions in data and simulated events, which can be observed before reweighting the simulated events, has vanished.

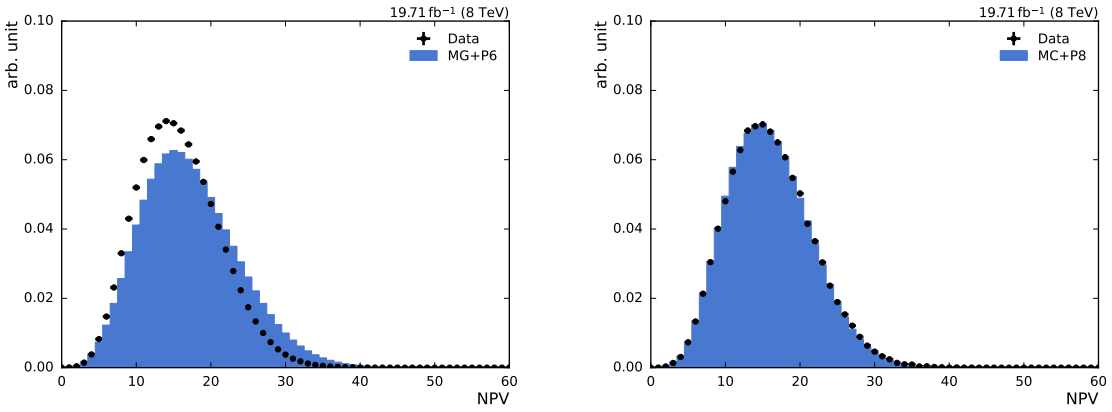

Figure 5.10: Number of reconstructed vertices in data and simulated events before (left) and after (right) the pileup reweighting.

5.5 Kinematic Distributions

The generated events are processed through the detector simulation and reconstruction in order to compare kinematic quantities of dijet events with simulated events on reconstructed level. Fig. 5.11 shows the transverse momentum p_{T}, the rapidity y and the azimuthal angle ϕ for the leading two jets. The p_{T} distribution is not that well described especially in the low p_{T} region. While the rapidity distributions agree within the tracker coverage of $|y| \leq 2.4$, there are larger discrepancies at higher rapidities.

Kinematic distributions of the dijet system are shown in Fig. 5.12. The azimuthal difference $\Delta\phi$ and the distance ΔR in the η-ϕ plane is well described by the simulated

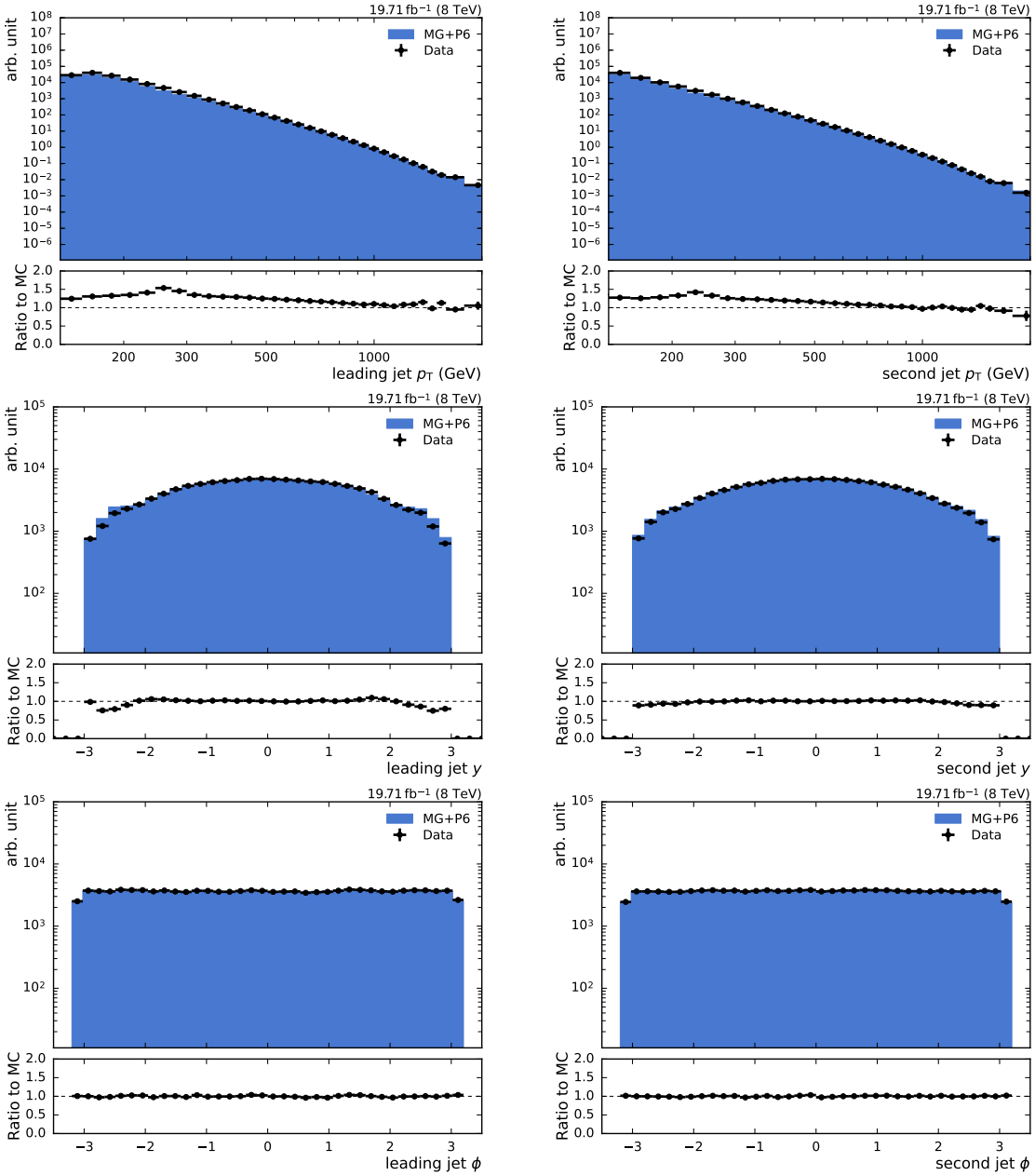

Figure 5.11: Kinematic quantities are shown for the leading (left) and the subleading jet (right) both for data (markers) and simulated events (histogram). The rows show the transverse momentum (top), the rapidity (middle), and the azimuthal angle of the jets (bottom).

events. The dijet mass $M_{1,2}$ and the average p_{T} of the dijet system are in agreement only at lower transverse momentum. At high dijet masses and transverse momenta, the theory significantly overestimates the data. The shape of the boost and rapidity separation distributions is well described for central jets. For very forward jets which are outside the coverage of the tracker, a significant difference is observed.

To illustrate the phase space origin of the dijet events in the various y^* and y_{b} bins, Fig. 5.13 shows the event yield as a function of the leading jet and second jet rapidities. The plot nicely illustrates, that all SS dijet events are contained in the bins with large values of y_{b} while all OS dijet events are collected in the bins with larger values of y^*.

5.5.1 Cross Section Comparison

The Monte Carlo simulation is used to compare the predictions including the detector simulation to the measured data distribution which is smeared by the finite detector resolution. Fig. 5.14 shows the prediction of the Madgraph and the Pythia 8 simulations as a ratio to the measured data spectrum. While the agreement in the inner detector region is fine, the shape and especially the normalization of the MC prediction cannot describe the data. Additionally, a fixed-order prediction of NLOJET++ is shown. This prediction on parton level is not corrected for non-perturbative effects. Still it describes the data, apart from the known non-perturbative and detector resolution effects, most accurately.

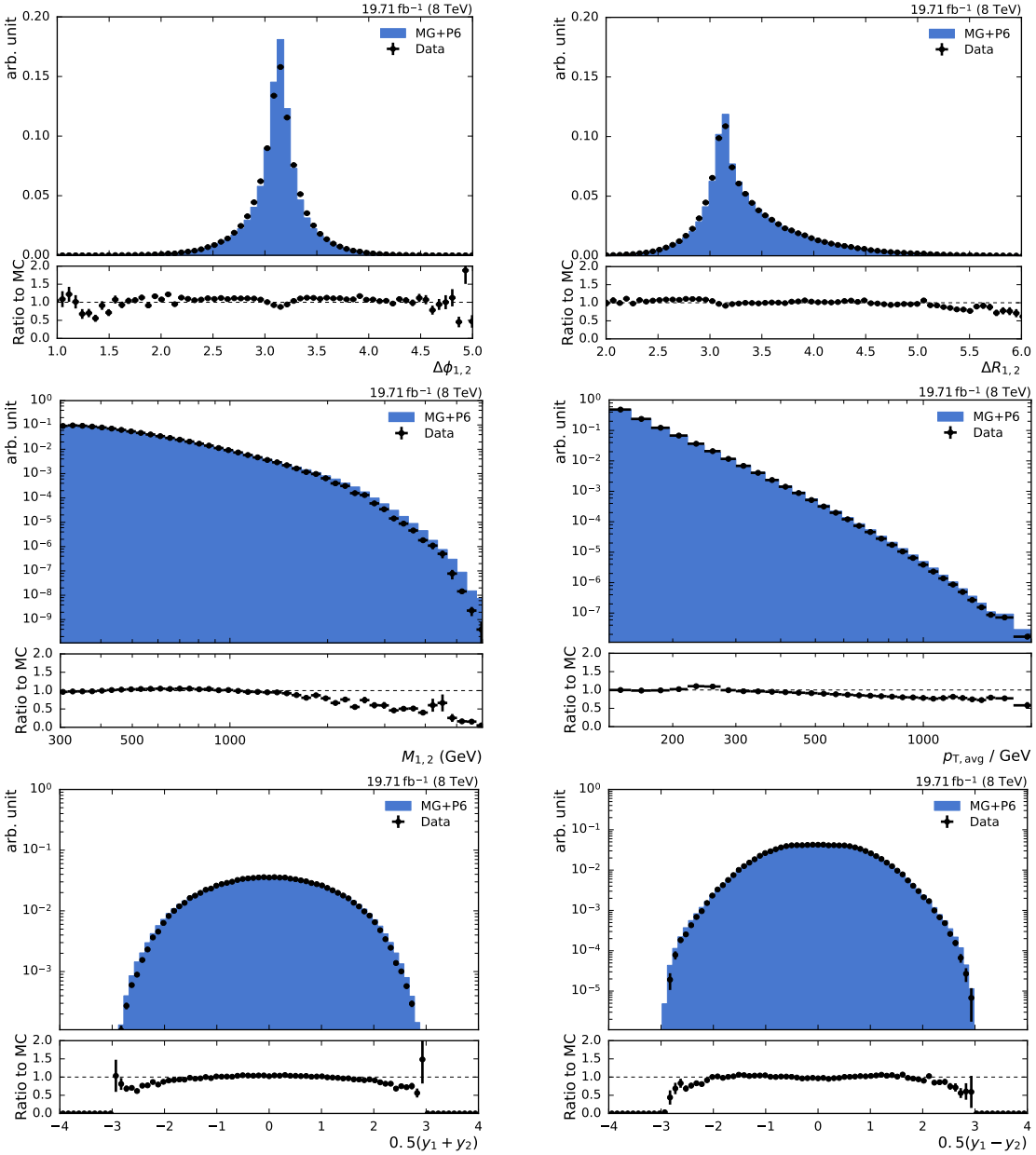

Figure 5.12: Kinematic quantities of the dijet system are shown for data (markers) and simulated events (histogram). The azimuthal separation $\Delta\phi_{1,2}$ and the distance in the ϕ-η plane $\Delta R_{1,2}$ are shown in the top row. The dijet mass $M_{1,2}$ and the average transverse momentum of the dijet system $p_{\mathrm{T,avg}}$ is shown in the middle row. The rapidity separation $0.5(y_1 - y_2)$ and the boost $0.5(y_1 + y_2)$ is shown in the bottom row.

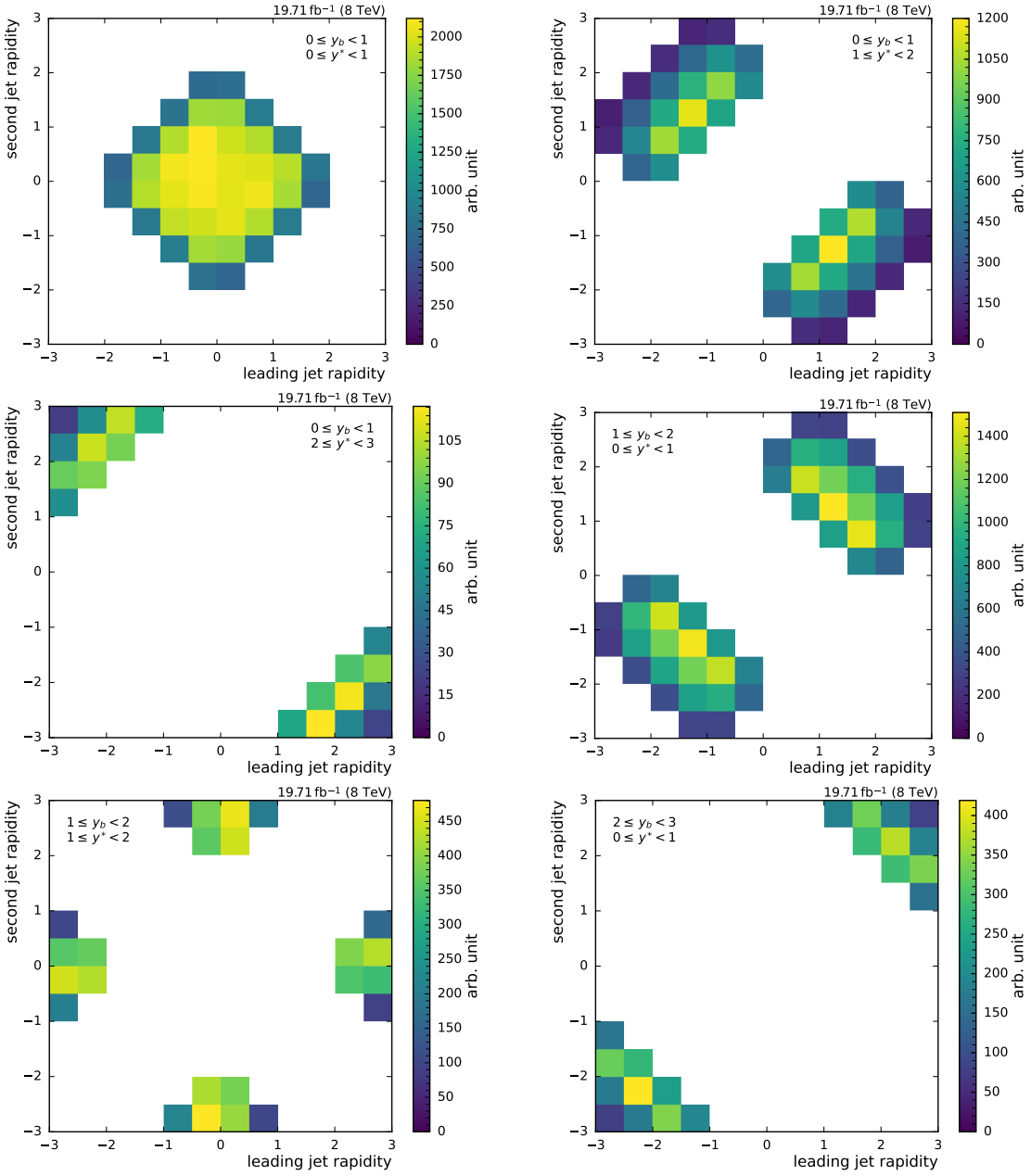

Figure 5.13: The distribution of the dijet events in the various y^* and y_b bins is shown as a function of leading jet and subleading jet rapidities. The distributions illustrate the separation of the dijet phase space in the various y^* and y_b bins.

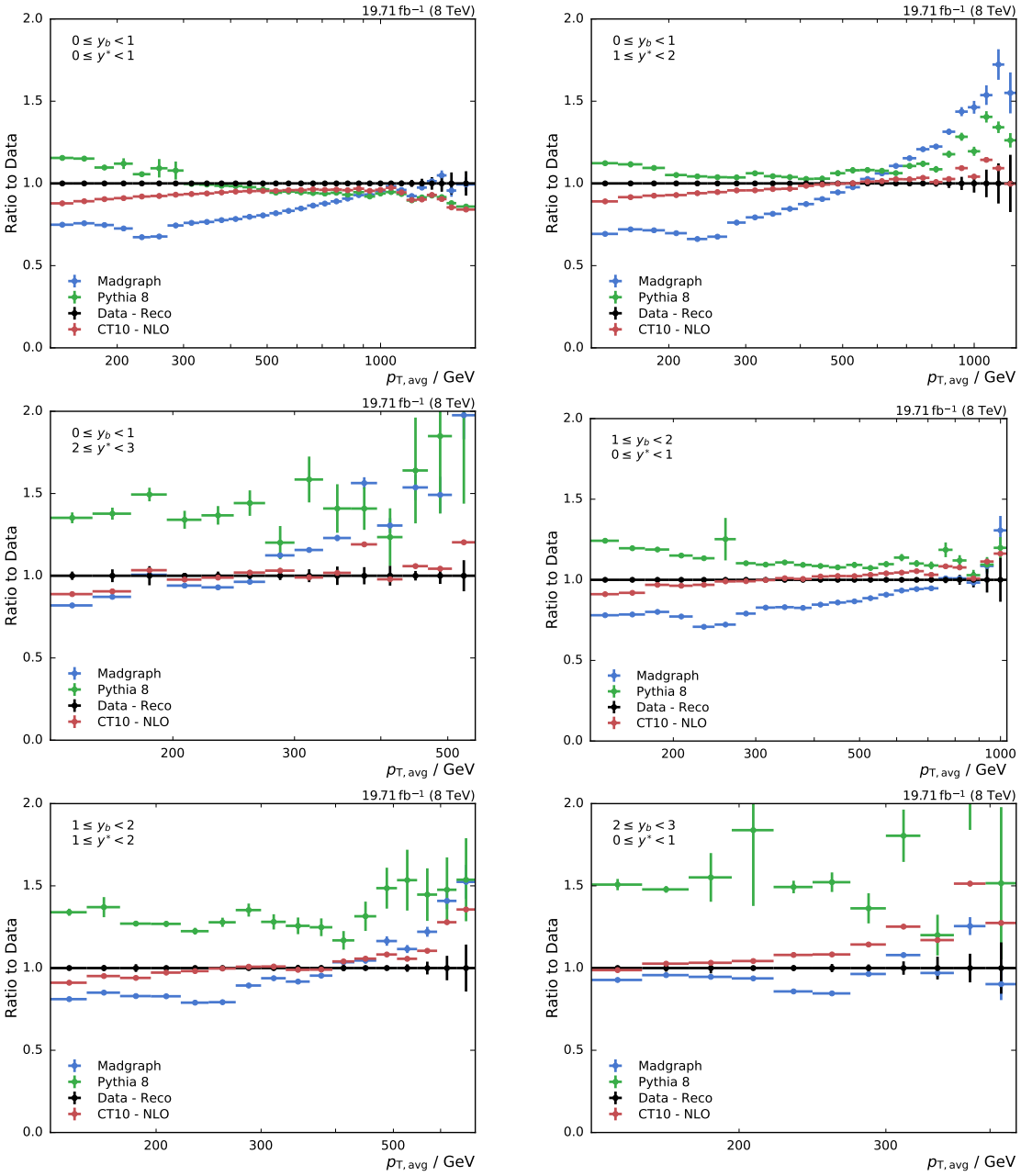

Figure 5.14: Comparison of the cross sections of data and simulated events of LO Monte Carlo generators at reconstructed level. The data distributions are normalized to the integrated luminosity, the simulated events to the number of generated events and the cross section. The ratio of the simulated events to the data points is shown.

5.6 Dijet Transverse Momentum Resolution

The transverse momentum of jets, which are measured in the CMS detector, is smeared because of the finite detector resolution. In order to correct the measured cross section for these detector effects, the momentum resolution of the observable has to be determined. As the simulated events are propagated through the detector simulation, the information at both particle level and reconstructed level is available. To compare particle-level jets and reconstructed jets, the jets belonging together have to be matched to each other. The distance ΔR in the η-ϕ plane between two jets is calculated as

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{5.4}$$

and needs to satisfy $\Delta R < 0.3$, roughly half of the jet size parameter 0.7. The jets closest to each other in the η-ϕ space are then matched. However, studies of the JERC working group [91] revealed that the resolution determined in simulated events is better than the resolution measured in data using data-based techniques. Therefore, the jet transverse momentum in simulated events is additionally smeared in order to match the resolution in data. Table 5.4 gives the smearing factor together with the assigned uncertainty indicated by the upwards and downwards variation. To consider the dependence on the detector geometry, the smearing factor is derived for various $|\eta|$-regions.

Table 5.4: The jet energy resolution in data is significantly worse than in simulated events. Therefore, the reconstructed jet transverse momentum in simulated events is smeared using a factor c to effectively match the resolution in data, following the recommendations of the JetMET group of CMS [91]. The uncertainty on the resolution is given by an upwards and downwards variation c_{up} and c_{down} of the smearing factor c.

	$\lvert\eta\rvert$						
	0.0 – 0.5	0.5 – 1.1	1.1 – 1.7	1.7 – 2.3	2.3 – 2.8	2.8 – 3.2	3.2 – 5.0
c							
nominal	1.079	1.099	1.121	1.208	1.254	1.395	1.056
downward	1.053	1.071	1.092	1.162	1.192	1.332	0.865
upward	1.105	1.127	1.150	1.254	1.316	1.458	1.247

The smearing of the reconstructed jet p_{T} is performed as a multiplicative scale factor based on the difference of $p_{\text{T}}^{\text{reco}}$ and $p_{\text{T}}^{\text{ptcl}}$, so that $p_{\text{T}}^{\text{reco}}$ is shifted to

$$p_{\text{T}}^{\text{reco}} = \max\left(0, p_{\text{T}}^{\text{ptcl}} + c(\eta) \cdot (p_{\text{T}}^{\text{reco}} - p_{\text{T}}^{\text{ptcl}})\right) \tag{5.5}$$

After the smearing, the response which is defined as

$$R = \frac{p_{\text{T,avg}}^{\text{reco}}}{p_{\text{T,avg}}^{\text{gen}}} \tag{5.6}$$

is calculated. Fig. 5.15 shows the response as a function of $p_{\text{T,avg}}^{\text{ptcl}}$ for each bin. As expected, the relative resolution of low-p_{T} jets is significantly worse than the one of high-p_{T} jets.

Since the response is dependent on both the detector region and the transverse momentum of the jets, the resolution is calculated as a function of $p_{\mathrm{T,avg}}^{\mathrm{ptcl}}$ separately in all y^* and y_{b} bin.

The resolution is then determined as the width of the distribution in each $p_{\mathrm{T,avg}}^{\mathrm{ptcl}}$ bin, as can be seen in Fig. 5.16. Both a Gaussian function as well as a double-sided Crystal Ball function have been studied in order to describe the observed behavior. Despite the large differences in the description of the non-Gaussian tails emphasized by the logarithmic representation, the width obtained from the fit of the distribution is very similar. Nonetheless, the Crystal Ball function is used to determine the resolution $\Delta p_{\mathrm{T,avg}}^{\mathrm{ptcl}} / p_{\mathrm{T,avg}}^{\mathrm{ptcl}}$ as it better describes the measured distributions, especially in the low-p_{T} region where the non-Gaussian tails are more pronounced.

Fig. 5.17 shows the resolution in each y^* and y_{b} bin as a function of $p_{\mathrm{T,avg}}$. The relative resolution is fitted using a modified version of the NSC formula.

$$\frac{\Delta_{p_{\mathrm{T,avg}}^{\mathrm{ptcl}}}}{p_{\mathrm{T,avg}}^{\mathrm{ptcl}}}(p_{\mathrm{T,avg}}^{\mathrm{ptcl}}) = \sqrt{\operatorname{sgn} N \left(\frac{N}{p_{\mathrm{T,avg}}^{\mathrm{ptcl}}}\right)^2 + \left(\frac{p_{\mathrm{T,avg}}^{\mathrm{ptcl}}}{\mathrm{GeV}}\right)^s \frac{S^2}{p_{\mathrm{T,avg}}^{\mathrm{ptcl}}} + C^2} \tag{5.7}$$

The formula is based on the calorimetric NSC formula which describes the resolution with terms for noise N, a stochastic component S and a constant shift C. Especially in the low-p_{T} region, in which the tracking has a non-negligible influence on the resolution due to the PF algorithm, a slightly better fit is obtained by using the modified resolution given in Formula 5.7. However, the influence of the different resolution formulas on the unfolded cross section is negligible as the spectrum starts at much higher $p_{\mathrm{T,avg}}$. Table 5.5 gives the parameters of the fit in each y^* and y_{b} bin of the measurement.

Table 5.5: Fitted parameters of the modified NSC formula characterizing the transverse momentum resolution in the y^* and y_{b} bins.

y_{b}	y^*	N	S	C	s
0 – 1	0 – 1	-2.68	1.43	0.03	-0.26
0 – 1	1 – 2	-8.00	5.81	0.04	-0.73
0 – 1	2 – 3	-0.02	1.98	-0.018	-0.36
1 – 2	0 – 1	-8.13	5.96	0.04	-0.73
1 – 2	1 – 2	2.85	1.16	0.02	-0.17
2 – 3	0 – 1	3.96	1.23	0.00	-0.18

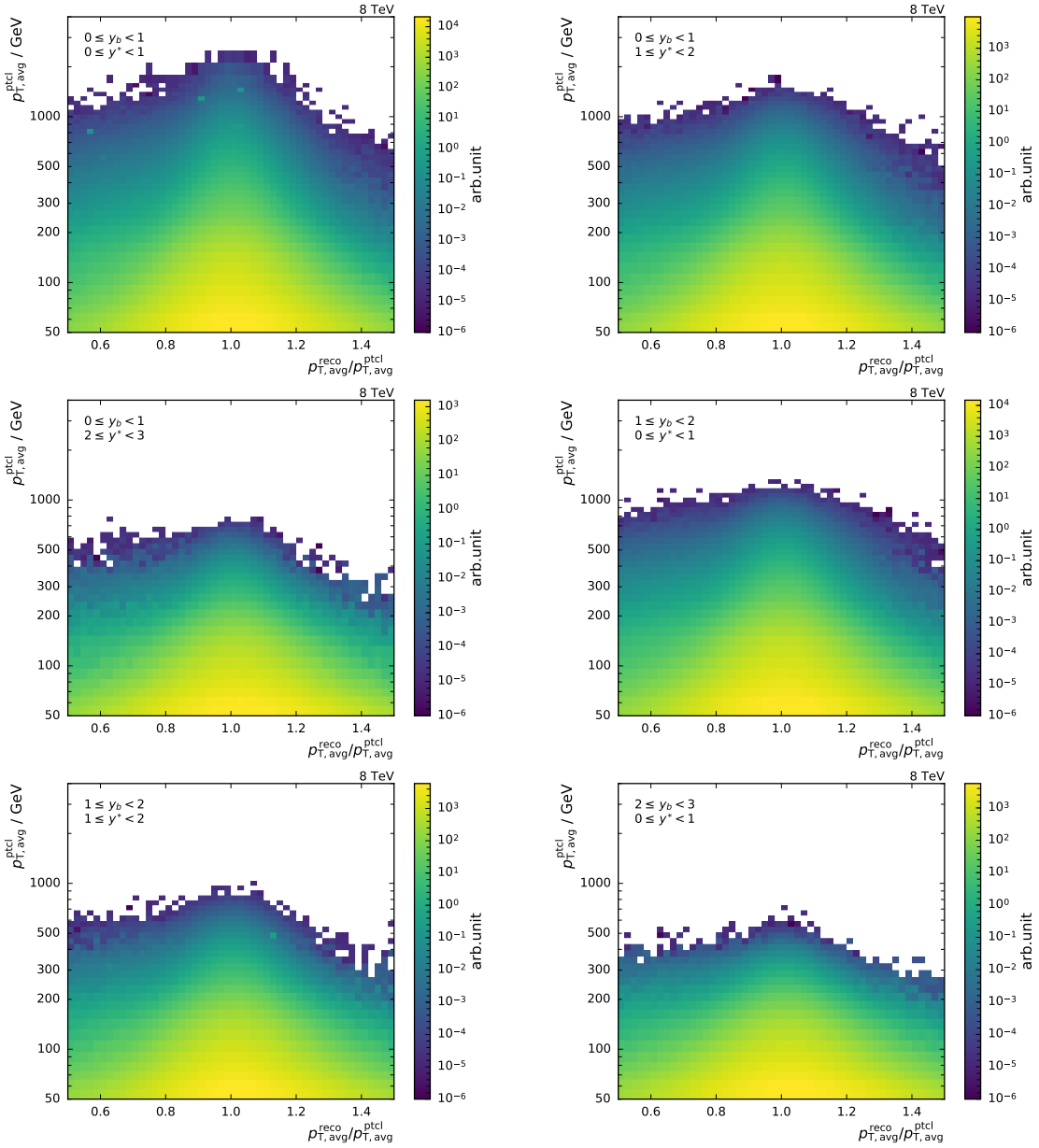

Figure 5.15: The response of reconstructed and particle-level jet transverse momentum as a function of the particle-level transverse momentum. The width of the distribution in each $p_{\mathrm{T,avg}}$ bin indicates the jet energy resolution, which improves from lower to higher values of $p_{\mathrm{T,avg}}$. The resolution is extracted separately for each bin and fitted using the NSC-formula.

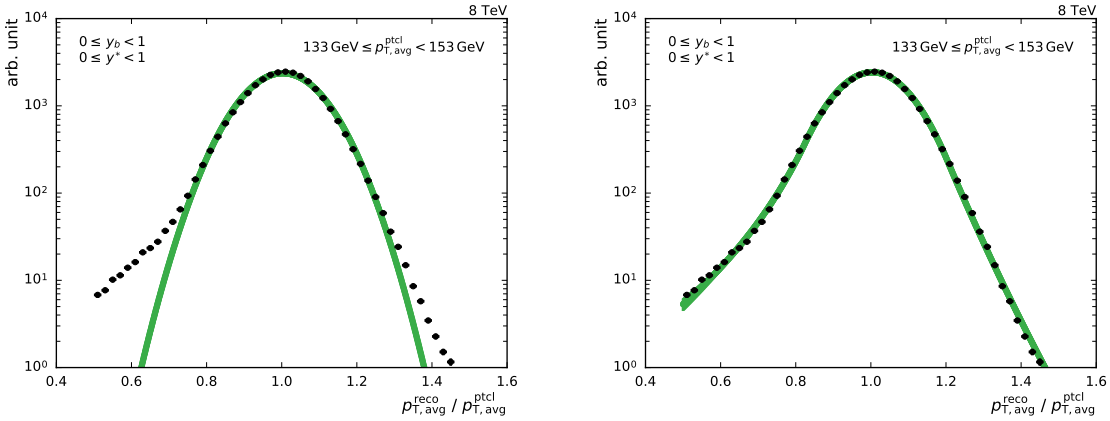

Figure 5.16: The resolution in each $p_{\text{T,avg}}$ bin is fitted using a Gaussian (left) or a double-sided Crystal Ball function (right). The fit using the Crystal Ball function better describes the non-Gaussian tails of the distribution. Therefore, it is favored in the determination of the jet energy resolution.

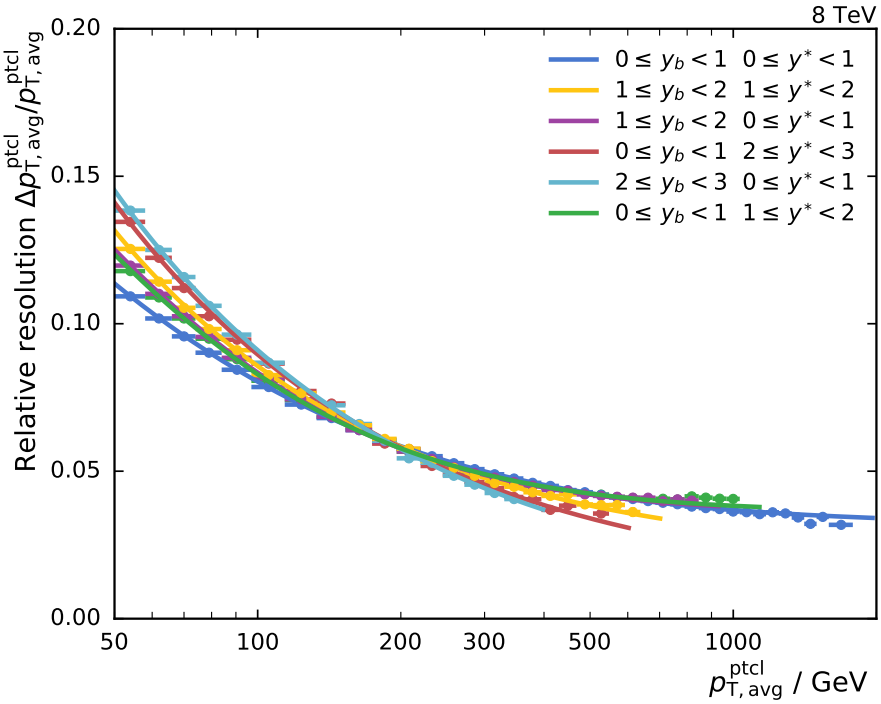

Figure 5.17: The jet energy resolution as a function of $p_{\text{T,avg}}^{\text{ptcl}}$ is shown for all y^* and y_{b} bins. The data points indicate the separate determinations while the solid lines give the results of the fits using the NSC-formula.

5.7 Unfolding of the Measurement

The final goal of this analysis is the comparison of the triple-differential dijet cross section measurement with higher-order QCD calculations on particle level. Due to finite detector acceptance and resolution, the jet transverse momentum is smeared causing differences between the particle-level $p_{\mathrm{T,avg}}^{\mathrm{ptcl}}$ spectrum and the measured spectrum $p_{\mathrm{T,avg}}^{\mathrm{reco}}$. To allow particle level comparisons, the measurement has to be unfolded. Thus, calculations from future Monte Carlo event generators can easily be compared with this measurement without the need to apply a detector simulation.

In this analysis, the iterative d'Agostini algorithm [92] is used, which is implemented in the RooUnfold [93] package. The unfolding process is regularized by the number of iteration steps in this algorithm. A higher number of iterations yields a reduced χ^2 but also increases the uncertainty and introduces larger bin-by-bin fluctuations and correlations. The regularization is optimized using simulated events and best results with low bin-by-bin correlations and low χ^2 are achieved using four iterations in the unfolding algorithm.

In principal, the response matrix for the unfolding algorithm can be populated directly using simulated events as they contain both the particle-level and reconstructed jets. However, this method has several drawbacks. The LO prediction does not describe the shape of the distribution in some phase space regions. Moreover, the limited number of events in the Monte Carlo sample, especially at high rapidities and high transverse momenta, introduces non-negligible statistical fluctuations in the response matrix. Because of these undesirable effects, an alternative strategy is employed to populate the response matrix.

Relying on the good description of the data spectrum by NLO predictions, a forward smearing technique is applied. The NLO prediction, obtained with the CT14-NLO PDF set and corrected for non-perturbative effects, is fitted using the function

$$f(p_{\mathrm{T,avg}}) = A_0 \left(\frac{p_{\mathrm{T,avg}}}{A_3}\right)^{-A_1} \left(1 - \frac{p_{\mathrm{T,avg}}}{A_3}\right)^{A_2}, \tag{5.8}$$

which describes both the normalization and shape of the distribution. Using toy Monte Carlo events, a flat $p_{\mathrm{T,avg}}$ spectrum weighted according to the fitted NLO distribution is generated. This distribution resembles the fastNLO cross section prediction and is used as the particle-level $p_{\mathrm{T,avg}}$ spectrum. All generated events are then smeared using the resolution function obtained in Sec. 5.6. The generation of these Monte Carlo toy events is very fast, and the response matrices can be filled with more than 100 million events each, resulting in negligible statistical fluctuations. Fig. 5.18 shows the response matrices used in the unfolding process. The matrices in the figure are normalized to the number of events in each particle-level bin to improve the readability. They are diagonal with small off-diagonal elements, which represent migrations between neighboring $p_{\mathrm{T,avg}}$ bins.

A closure test was performed by unfolding the generated smeared spectrum in two ways. Once, the smeared spectrum was obtained using the same events which were used to fill the response matrix and once, the spectrum was generated using statistically independent events. In the first case, exactly the same truth spectrum should be re-obtained after

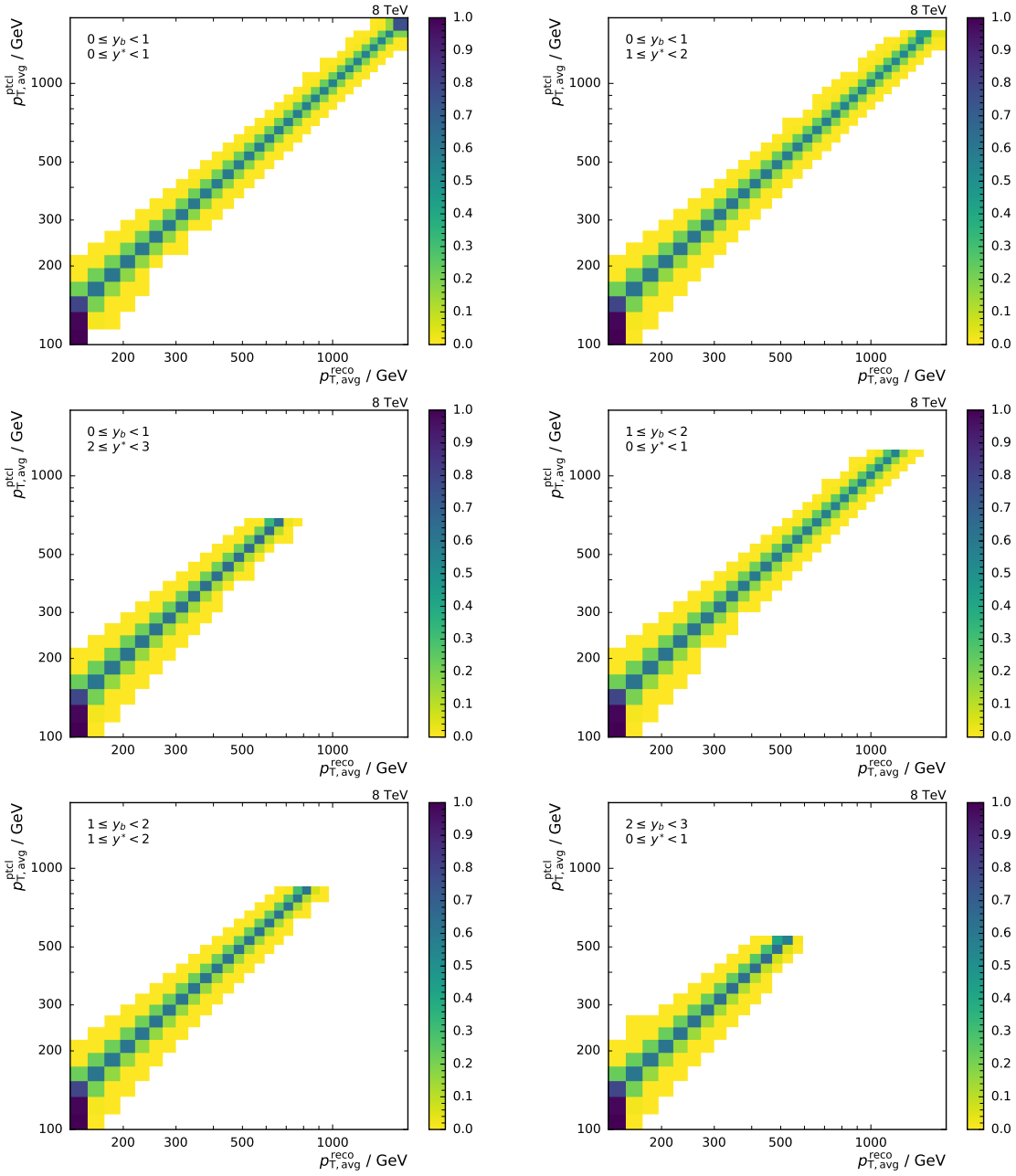

Figure 5.18: Response matrices illustrating the bin migrations. They are obtained using the forward smearing method and are normalized to the number of events of particle-level events to improve readability. The response matrices are diagonal with small off-diagonal entries indicating bin migrations between neighboring $p_{\mathrm{T,avg}}$ bins.

unfolding, while in the second case, the results should agree within statistical uncertainties. Fig. 5.19 demonstrates that both assumptions hold for the applied unfolding technique.

Another closure test was performed by filling the response matrices directly from the simulated events of the Madgraph data sample and comparing the unfolded cross section with the one obtained using the forward smearing technique. In both cases, the results are consistent within uncertainties. However, the statistical uncertainties are very large due to the limited number of events in the Madgraph sample.

The applied unfolding technique does not consider fluctuations between y^* and y_b bins. However, migrations between neighboring y^* and y_b bins were studied and found to be very small. Therefore, it is justified to perform the unfolding in each y^* and y_b bin separately.

Statistical uncertainties of the data distributions are propagated through the unfolding procedure using toy experiments. The procedure as well as its results are discussed in detail in Sec. 5.8.2 when experimental sources of uncertainties are discussed. The unfolded cross sections are shown in Sec. 5.9, in which the comparison to NLO calculations is presented.

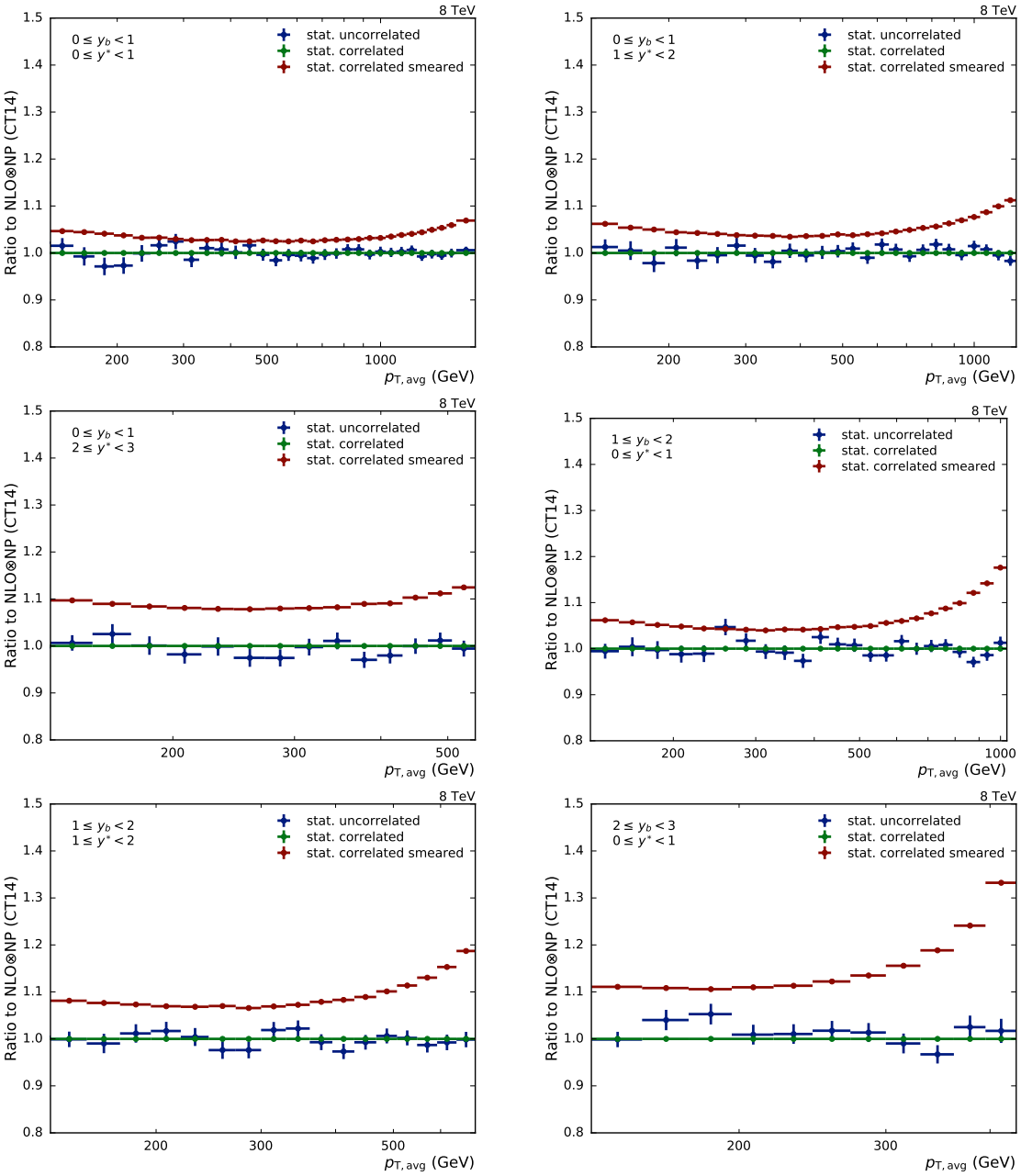

Figure 5.19: Closure test of the employed unfolding procedure. Once, the smeared spectrum is obtained from the same events that were used to fill the response matrix and once from statistically independent events. As expected, unfolding the former spectrum (green line) again yields exactly the truth spectrum. Unfolding the statistically independently smeared spectrum gives compatible results within statistical uncertainties (blue line).

5.8 Experimental Uncertainties

In this section all experimental uncertainties which affect the cross section measurement are discussed: The statistical uncertainties, jet energy scale and resolution uncertainties, the luminosity uncertainty, and a residual uncertainty accounting for systematic errors due to the jet ID and trigger efficiencies.

An overview depicting all experimental uncertainties can be found in Fig. 5.22 at the end of the section. It presents all sources of experimental uncertainty in combination with the total experimental uncertainty, obtained by adding all individual sources in quadrature. The total uncertainty amounts to 4 % in the measurement bins involving jets at central rapidity and increases up to 25 % in bins where jets from worse understood phase space regions contribute, see Table 5.6.

Table 5.6: Five experimental sources of uncertainty affect the accuracy of the measurement. The dominant source arises from jet energy scale corrections. The total uncertainties are as small as 5% in the best understood region and increase up to 25% in worst understood phase space regions.

Uncertainty source	Min	Average	Max
Jet energy scale	2.5 %	4.0 %	11 %
Luminosity	2.6 %	2.6 %	2.6 %
Jet energy resolution	0.3 %	1 %	2.5 %
Statistical	0.1 %	1.5 %	22 %
Residual	1 %	1 %	1 %
Total	4 %	6 %	25 %

5.8.1 Uncertainty on Luminosity Measurement

As discussed in Sec. 3.2, the luminosity is measured from the number of clusters in the silicon pixel detector. The uncertainty on the luminosity measurement for the 2012 LHC run is estimated to be 2.5% (syst.) and 0.5% (stat.) [46]. As the luminosity uncertainty translates into a normalization uncertainty on any absolute cross section measurement, a combined systematic uncertainty of 2.6% is assigned, which is fully correlated across all bins.

5.8.2 Unfolding and Statistical Uncertainties

Statistical uncertainties of data points are propagated through the unfolding using a toy MC technique in which the data are smeared according to their statistical uncertainties. Then, the unfolding procedure is repeated using the smeared spectrum. One million of such toy spectra are used to propagate the statistical uncertainty. Fig. 5.20 shows the relative statistical uncertainty before and after the unfolding procedure. The uncertainty slightly increases during the unfolding process.

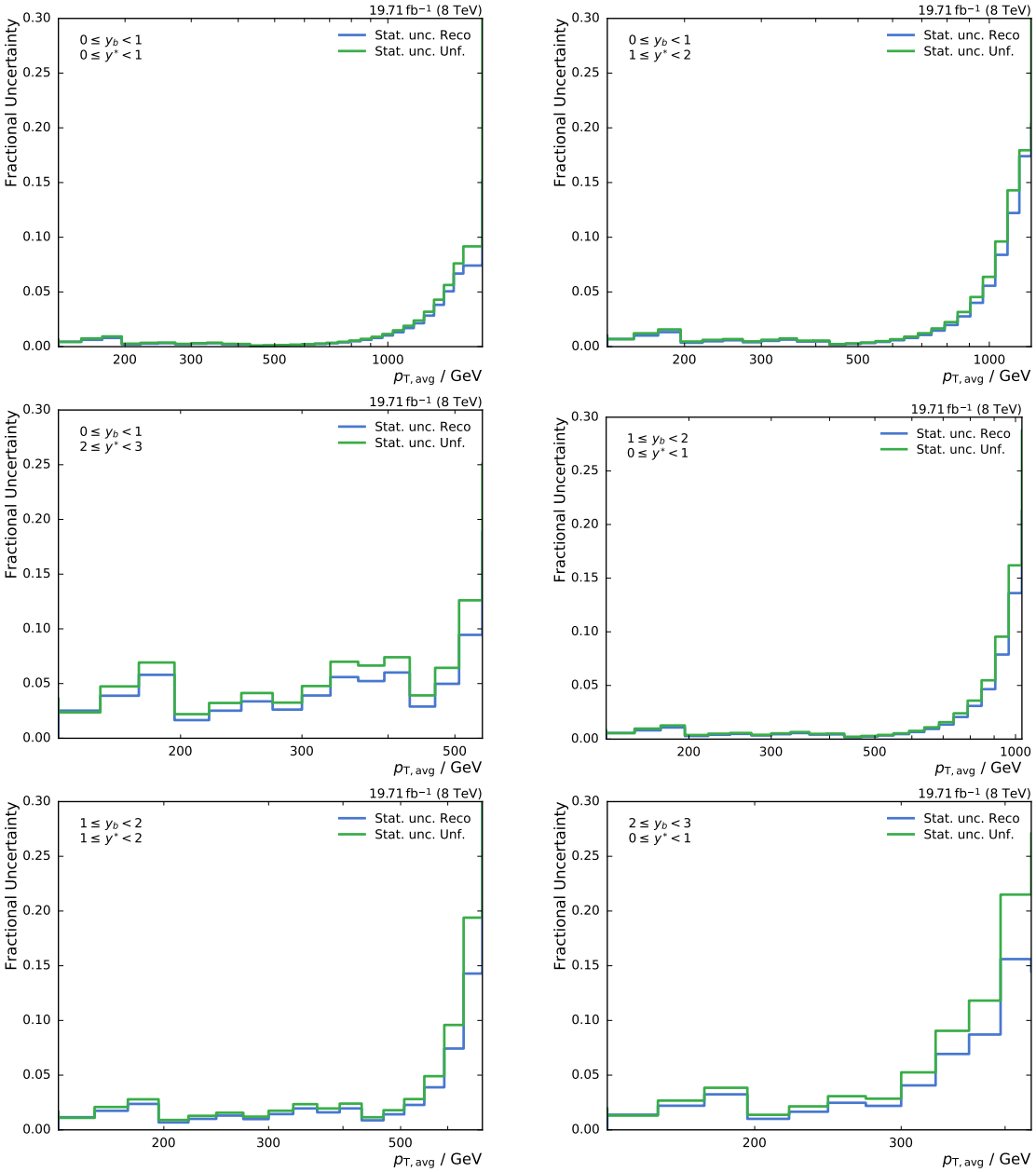

Figure 5.20: The statistical uncertainties of the measured and the unfolded data. Depending on the unfolding procedure, the uncertainties can slightly increase.

Furthermore, the unfolding introduces a correlation between bins due to event migrations. These correlations are significant for neighboring bins in p_{T} and negligible between bins far apart in the phase space. Fig. 5.21 shows the correlations of the statistical uncertainty after the unfolding. Of course, these correlations have to be taken into account in any statistical analysis of the data such as a PDF fit.

5.8.3 Jet Energy Correction Uncertainties

The dominant part of the experimental uncertainties derives from the jet energy calibration, which corrects the measured jet energy for a variety of detector effects and is discussed in Sec. 3.5.4. As the corrections are afflicted with multiple sources of systematic uncertainty, all sources are propagated individually to the cross section to preserve their correlations.

The JEC uncertainties are split into 25 mutually independent sources of uncertainty, in which each source is fully correlated in p_{T} and η, but uncorrelated to all other sources, and presents a 1σ shift. As these uncertainties can be asymmetric, the upwards and downwards variation of each source are treated separately. The sum in quadrature of all sources yields the total JEC uncertainty. Therefore, they can be treated in exactly the same way as the PDF eigenvector sets, which were discussed in Sec. 4.4.2. The sources of uncertainty are grouped in four categories according to their origin. In the following, a short summary of the sources is given. More details about the jet energy corrections and uncertainties are given in [81]. The Figs. A.2–A.7 in the appendix show the size of each of the 25 sources separately.

Pileup JES Differences in the transverse momentum between the true offset and the random cone offset are observed in simulated events. This difference is propagated using Z/γ+jet and dijet balancing methods to estimate the residual pileup uncertainty after the calibration.

Relative JES The relative η-dependent corrections calibrate forward jets using balanced dijet events. The largest contribution to the uncertainty arises from jet energy resolution and soft radiation bias corrections.

Absolute JES The absolute calibration of the jet energy scale relies on Z/γ+jet and multi-jet events. The uncertainties are related to the lepton momentum scale for muon and the single pion response in the HCAL. Observed differences in applied methods can be traced back to neutrinos and ISR and are accounted for in these sources of uncertainty.

Flavor JES Differences in the flavor response are studied using simulation by cross-checking the results with quark- and gluon-tagged γ+jet and Z+jet events. The uncertainty is derived based on differences observed in the Pythia 6 and Herwig++ simulation.

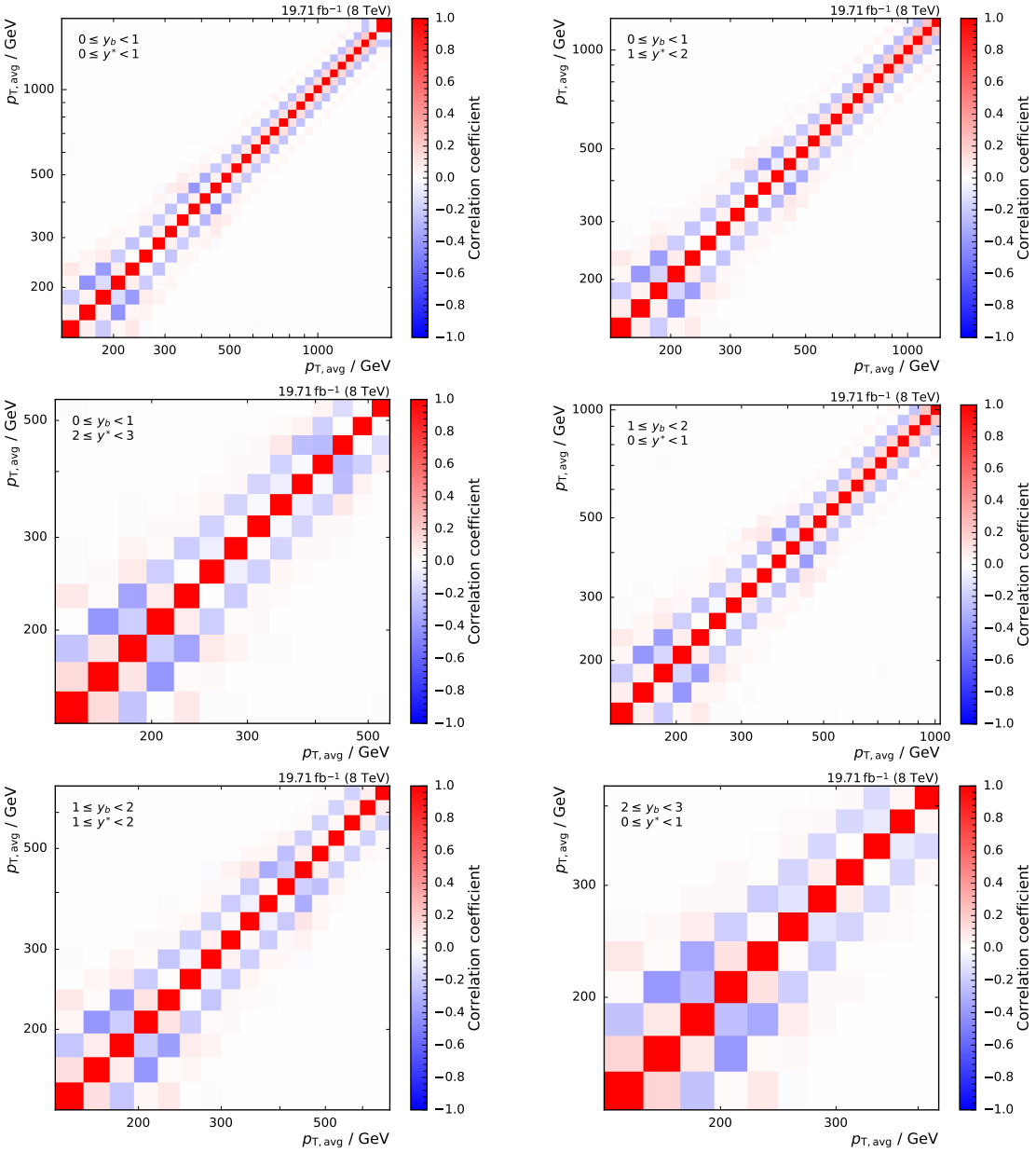

Figure 5.21: Correlations of the statistical uncertainty introduced by the unfolding procedure. Neighboring bins exhibit a significant correlation or anti-correlation through bin migrations.

5.8.4 Jet Energy Resolution Uncertainty

The jet energy resolution, which was derived in Sec. 5.6, is used to populate the response matrix using a forward smearing technique in the unfolding procedure. Therefore, a dependence of the unfolded cross section on the jet energy resolution is introduced.

Table 5.4 shows the smearing factors, which were applied on reconstructed simulated events to obtain the actual resolution in data. The official recommendations include offset variations of these smearing factors to estimate the uncertainty on the resolution and are also given in Table 5.4. The determination of the resolution is repeated with the applied upwards and downwards variation of the resolution smearing factor. The unfolding procedure is also reiterated using the variations of the resolution, and the differences of the obtained cross section to the nominal cross section are accounted for with a systematic uncertainty. The influence on the cross section is comparably small, about 1% in low y^* and y_b and increasing up to 3% for the highest y^* value.

An overview of all uncertainties is given in Fig. 5.22, which shows their relative size in each bin.

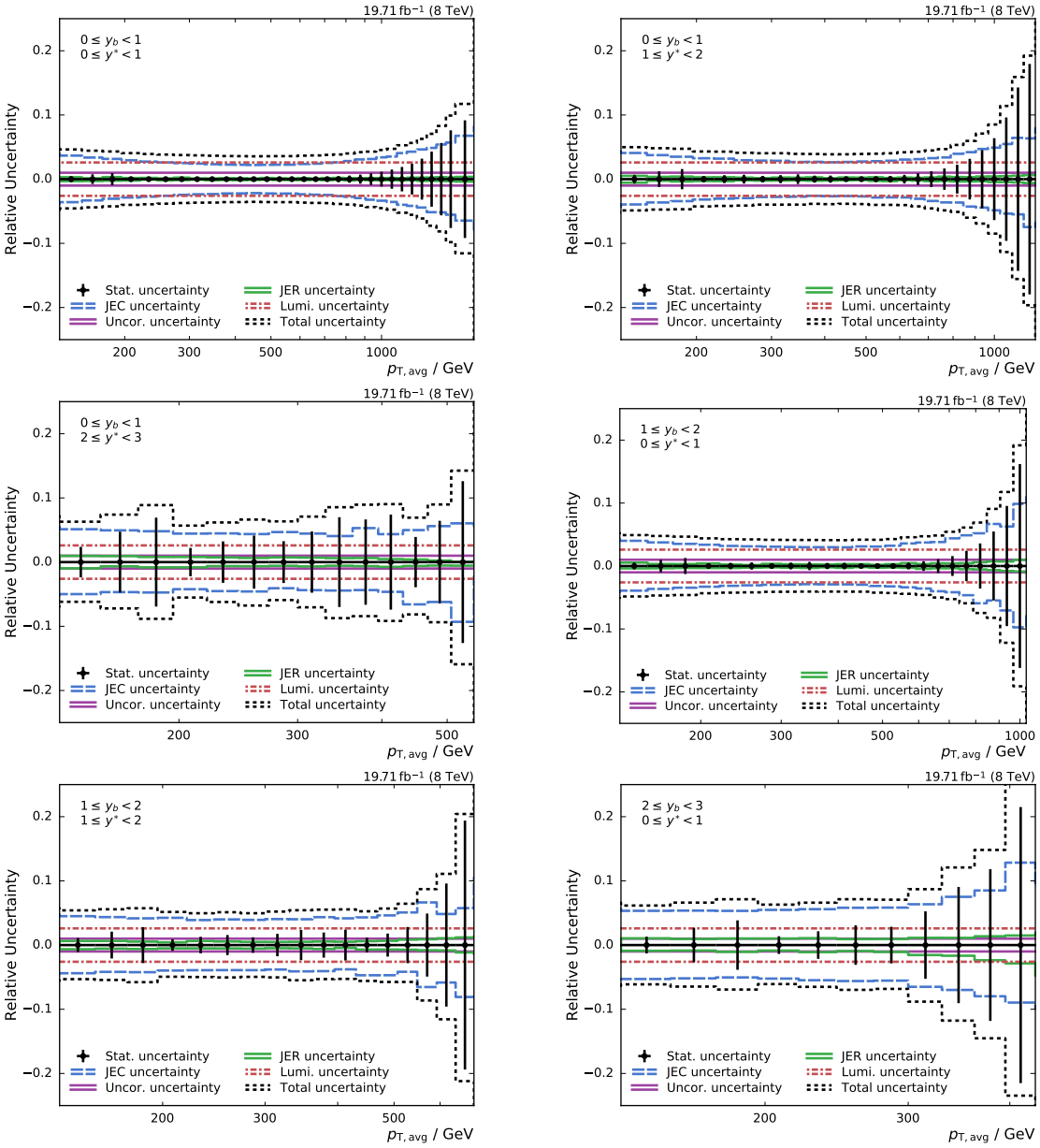

Figure 5.22: Overview of all experimental uncertainties affecting the cross section measurement. The error bars indicate the statistical uncertainty after unfolding. The colored lines give the uncertainties resulting from jet energy corrections, jet energy resolution, luminosity and residual effects. The total uncertainty is yielded by adding the individual sources of uncertainty in quadrature.

5.9 Comparison with NLO Predictions

After unfolding the measurement, it is finally possible to compare the measured cross sections with the NLO calculations obtained in Sec. 4. A general comparison is given in Fig. 5.23, which shows the data overlaid with the NLOJET++ prediction obtained with the NNPDF 3.0 PDF set. The fixed order NLO calculation is corrected for non-perturbative effects. The measurement and the NLO predictions agree over many orders of magnitude of the cross section.

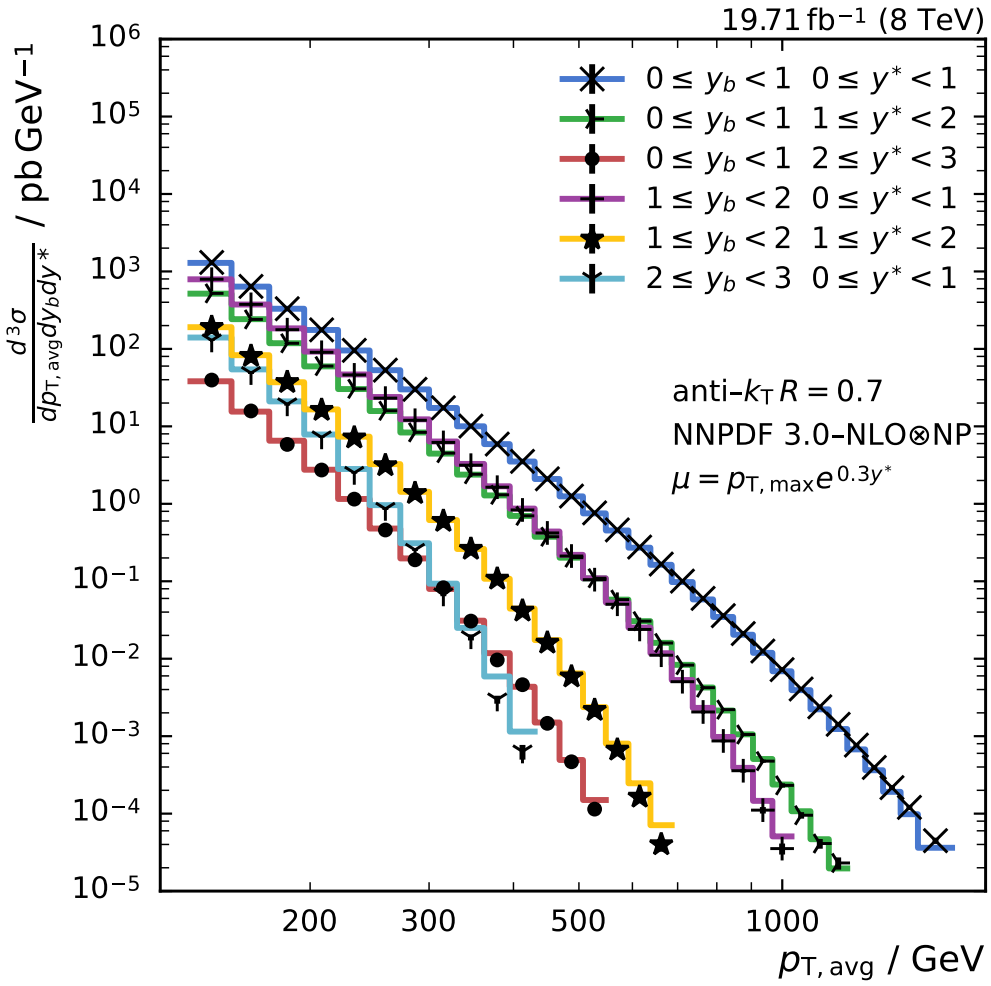

Figure 5.23: The triple-differential dijet cross section in six bins of y^* and y_b. The data are indicated by different markers for each bin and the theory obtained with NLOJET++ and NNPDF 3.0 is depicted by colored lines. Apart from the boosted region, the data is well described by NLO theory calculations over many orders of magnitude.

A more detailed comparison is possible when the ratio of data to theory is calculated. Fig 5.25 presents such ratios for calculations using different PDF sets. Apart from a few

phase space regions, the predictions using all PDF sets, except for the ABM 11 PDF set well describe the data. The predictions using the ABM 11 PDF set systematically underestimate the data. This behavior is well known, e. g. from [16, 52], and can be traced back to a soft gluon PDF accompanied with a low value of $\alpha_s(M_Z)$ in the PDF.

In the central region, at low y_b and y^*, a systematic difference of up to 20 % between data and NLO predictions is observed for transverse momenta beyond 1 TeV. This mismatch is presumably caused by missing electroweak corrections, which are positive and sizable at low rapidity and p_T larger than 1 TeV [16, 94]. Theory colleagues are currently working on providing the electroweak corrections specifically for this measurement, but the corrections were not available in time.

Figure 5.24 shows ratios of the data to the predictions of the NLOJET++, Powheg, and Herwig 7 NLO MC generators. There are significant differences observed between the predictions of the studied MC generators. These could originate from the PDFs or the scale choices in each MC generator which are not trivially adaptable. In general, Herwig 7 better describes the data in the central region while Powheg prevails in the forward region.

Especially phase space regions in which the data discriminate between the predictions of different PDF sets are interesting for PDF studies. As discussed in Sec. 4.1, the bins involving boosted dijet events, in which the accessed fractional proton momenta x_1 and x_2 are very different, are predestined for PDF studies. The predictions of the different PDF sets, especially compared to the ones of the NNPDF set, yield different results in the boosted region and are afflicted with large PDF uncertainties. Moreover, none of the investigated PDF sets yields a good description of the data in this phase space region, see Fig. 5.24 bottom right, making it a very interesting subject for PDF studies.

5.10 Summary

In this chapter, the complete data analysis of the triple-differential dijet cross section measurement has been presented. The reconstructed spectrum has been unfolded with the iterative d'Agostini algorithm to correct for all detector dependent effects. By comparing with NLO pQCD calculations, it was found that the data are well described in most of the studied phase space. Systematic deviations at some edges of the investigated phase space were observed and probably could be traced back to missing electroweak corrections and the PDFs. The high experimental precision of the data and the comprehensive study of all sources of uncertainties including their correlations allows to include this measurement in a PDF fit to derive constraints.

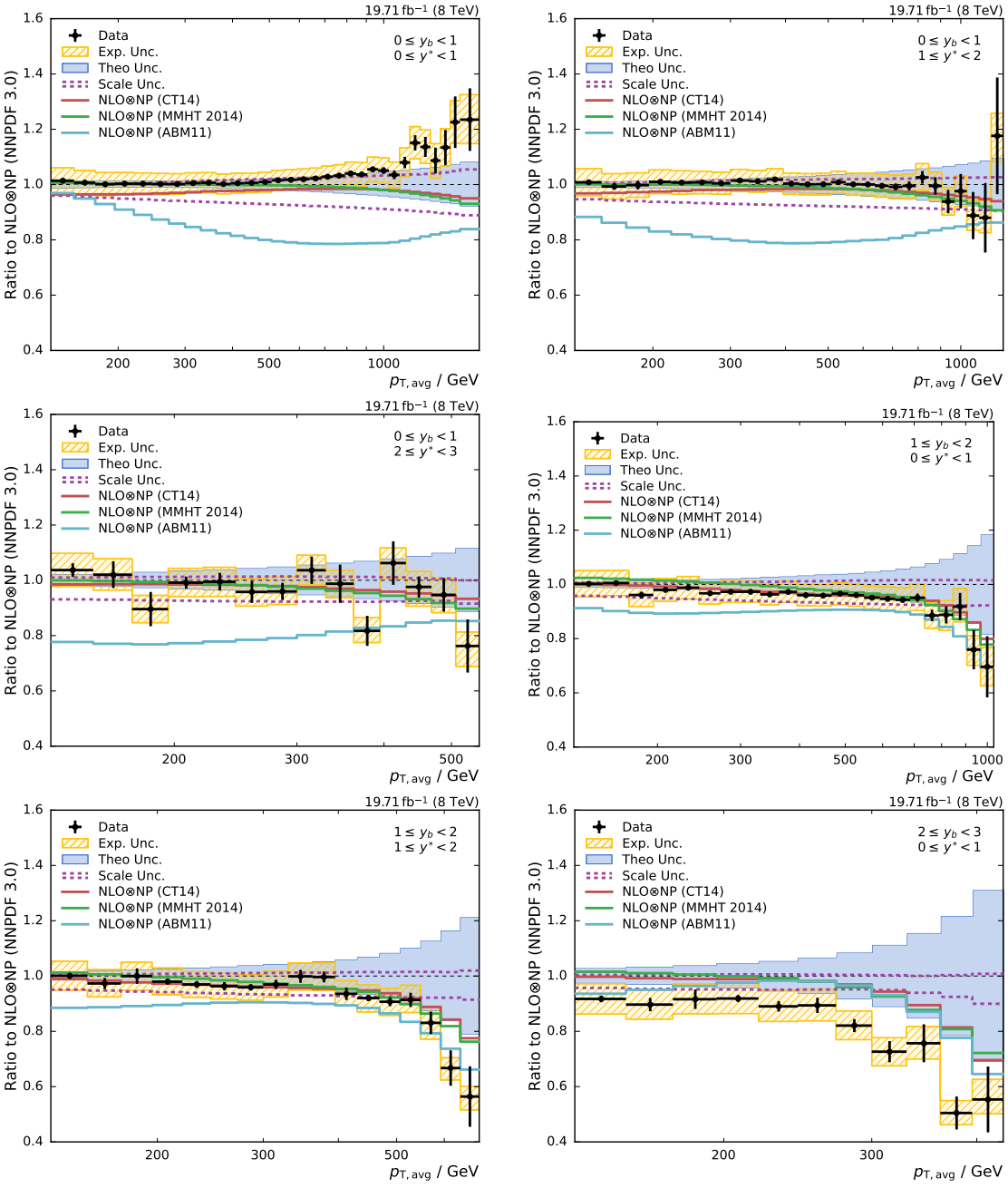

Figure 5.24: Ratio of the triple-differential dijet cross section to the NLOJET++ prediction ($\mu = p_{\mathrm{T,max}} e^{0.3y^*}$) using the NNPDF 3.0 set. The data points including statistical uncertainties are indicated by markers, the total experimental uncertainty is represented by the hatched band. The blue band shows the PDF and NP uncertainties quadratically added and the continuous colored lines give the predictions calculated with different PDF sets.

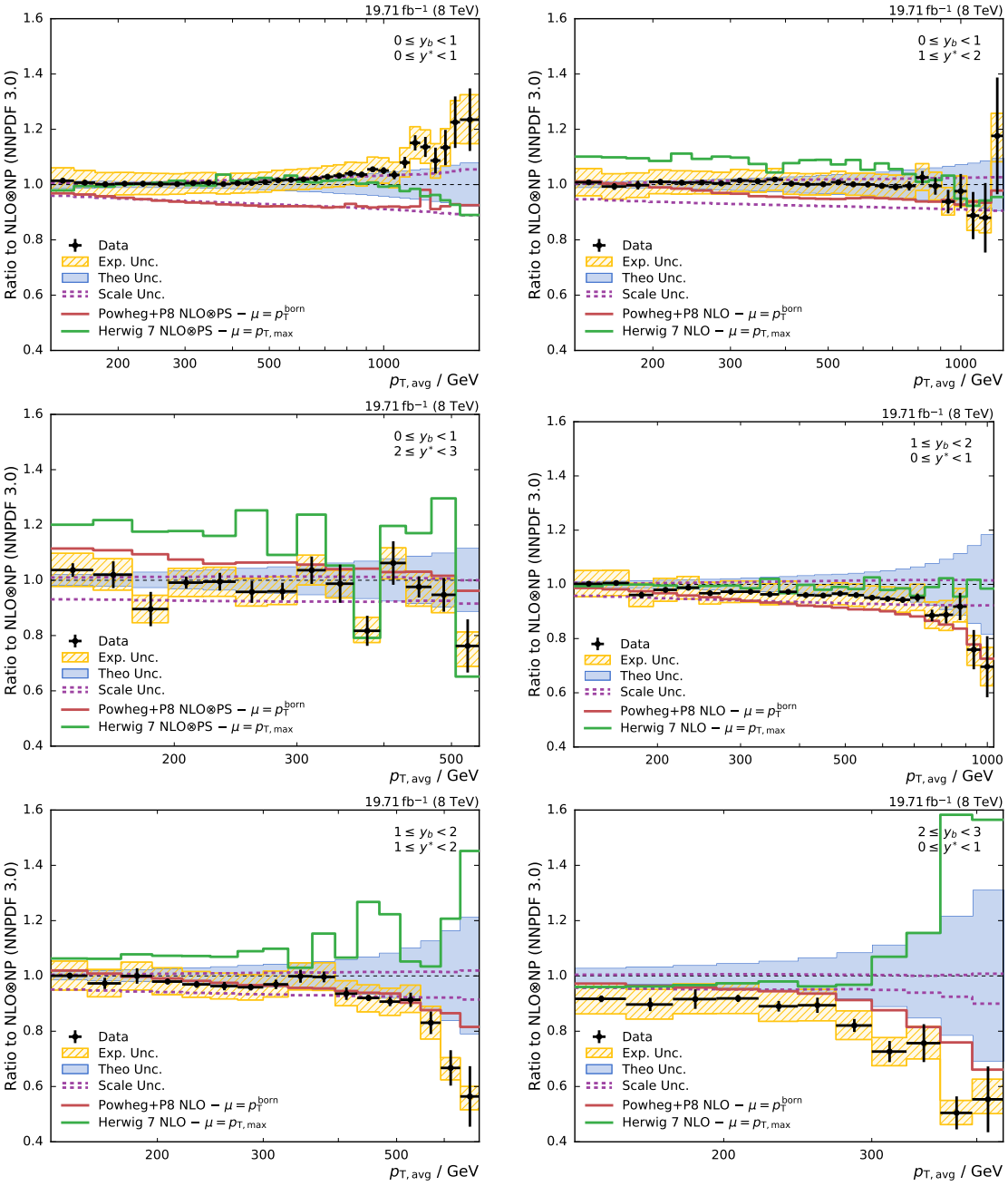

Figure 5.25: Ratio of the triple-differential dijet cross sections to the NLOJET++ prediction ($\mu = p_{T,max} e^{0.3y^*}$) using the NNPDF 3.0 set. The data points including statistical uncertainties are indicated by markers, the total experimental uncertainty is represented by the hatched band. The blue band shows the PDF and NP uncertainties quadratically added. The predictions of the NLO MC generators Powheg+P8 and Herwig 7 are depicted by the red and green lines, respectively.

CHAPTER 6

PDF Constraints and Determination of the Strong Coupling Constant

The proton PDFs are an essential ingredient in almost all theory predictions at the LHC. Since the PDFs cannot be calculated in perturbative QCD, they are determined from fits to experimental data of collider and fixed-target experiments. The precise DIS data from the combined measurements of the HERA-I and HERA-II run periods[1] provides the base data set since it covers a large range in x and Q of the kinematic phase space. By including more data from different experiments, which provide constraints in additional phase space regions, the precision of the PDFs can be improved. In this chapter, the impact of the triple-differential dijet cross section on the PDFs is studied by performing a fit of the combined HERA DIS data and the triple-differential dijet cross section measurement.

As discussed in Sec. 5.9, the electroweak corrections have not been available in time for this thesis. They are large only at highest transverse momenta and central rapiditities. Thus, the bins with $p_{\mathrm{T,avg}} > 1000\,\mathrm{GeV}$ have not been considered in the PDF studies removing 11 out of 144 bins. However, the experimental uncertainties of the neglected bins are comparably large. Hence, their impact on the fit result is small.

6.1 Correlation between Dijet Cross Section and PDFs

To determine the phase space regions in which the triple-differential dijet cross section measurement is sensitive to the PDFs, the correlation between the cross section $\sigma(\mu)$ and the PDFs $xf(x,\mu^2)$ for any parton flavor f is calculated. The PDF sets of the NNPDF collaboration are an ensemble of replicas i which sample variations in the PDF parameter space within the PDF uncertainties. The correlation coefficient $\varrho_f(x,\mu)$ between the cross section and the PDF for flavor f at a point (x,μ) is calculated by evaluating the mean and standard deviation from an ensemble of N replicas as

$$\varrho_f(x,\mu) = \frac{N}{(N-1)} \frac{\langle \sigma(\mu)_i \cdot xf(x,\mu^2)_i \rangle - \langle \sigma(\mu)_i \rangle \cdot \langle xf(x,\mu^2)_i \rangle}{\Delta_{\sigma(\mu)} \Delta_{xf(x,\mu^2)}} \tag{6.1}$$

[1] In the following, the combined HERA-I and HERA-II DIS data sets are always referred to as HERA DIS data.

where $\Delta_{\sigma(\mu)}$ and $\Delta_{xf(x,\mu^2)}$ are the standard deviations of the dijet cross section prediction and of the PDF with flavor f, respectively. Fig. 6.1 presents the correlation coefficient between the dijet cross section and the gluon, u valence quark, and d valence quark PDFs for the central region ($y_\mathrm{b} < 1, y^* < 1$) and the boosted region ($2 \leq y_\mathrm{b} < 3$).

The correlation between the gluon PDF and the dijet cross section is large in the central region for $p_{\mathrm{T,avg}} > 1\,\mathrm{TeV}$ and momentum fractions $0.1 < x < 0.5$. In the boosted region there is a large correlation for $200\,\mathrm{GeV} \leq p_{\mathrm{T,avg}} < 400\,\mathrm{GeV}$ and momentum fractions $0.2 < x < 0.7$. In contrast, the correlation between the dijet cross section and the u valence and d valence quark PDFs is much smaller, especially in the central region. However, the correlation in the boosted region is more pronounced, particularly for momentum fractions $x > 0.5$. Additional figures with the correlations for the other bins can be found in the appendix (Figs. A.10–A.13). Based on the results of the correlation studies, a significant impact on the PDFs at high x is expected by including the dijet cross sections in a PDF fit.

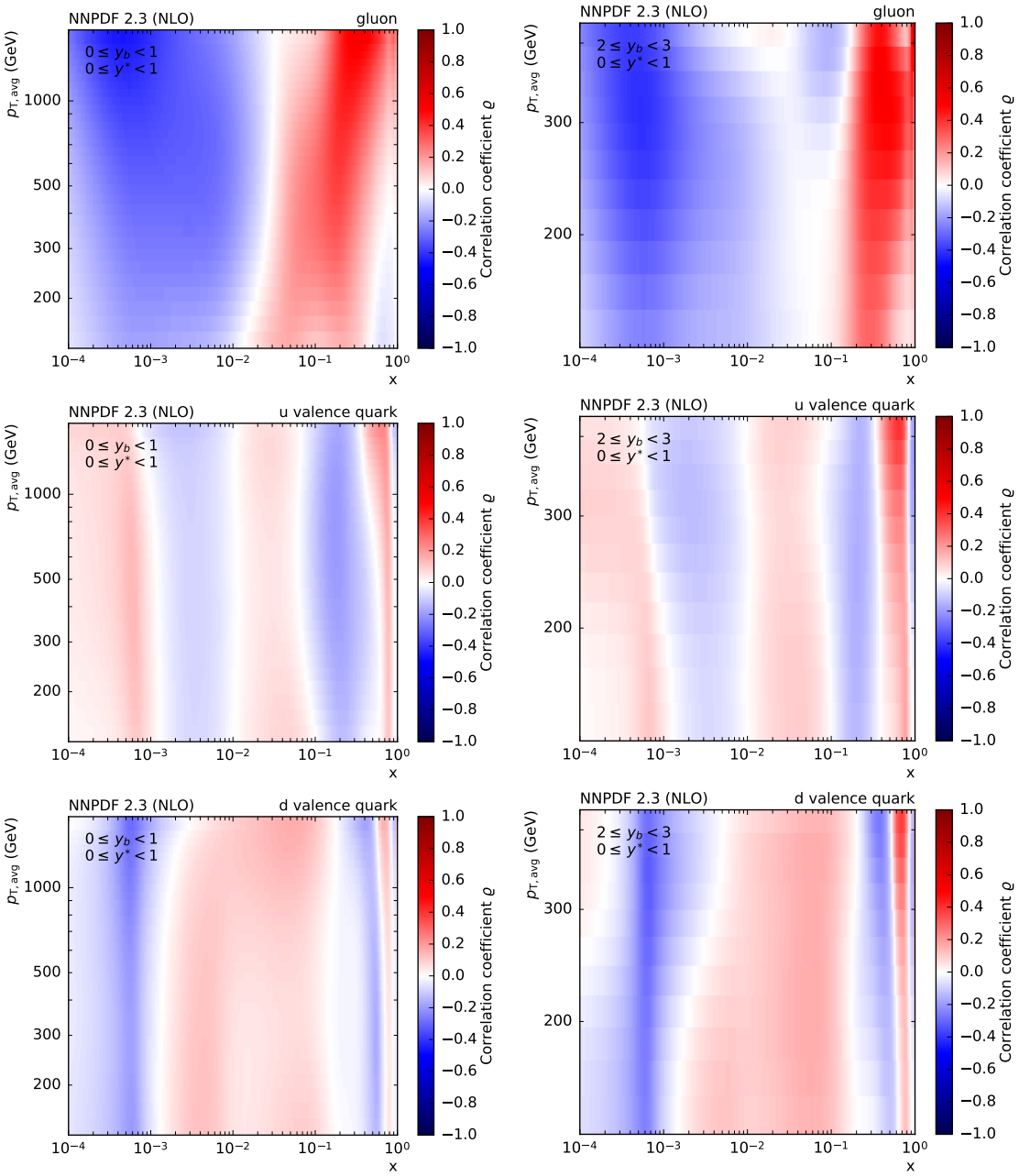

Figure 6.1: The correlation coefficient between the triple-differential dijet cross section and the gluon (top row), u valence quark (middle row), and d valence quark PDFs (bottom row) as a function of the momentum fraction x and the energy scale μ of the hard process. The correlation is shown for the central bin $y_\mathrm{b} < 1$, $y^* < 1$ (left) and for the boosted region $2 \leq y_\mathrm{b} < 3$ (right).

6.2 The XFitter Framework

The constraints of the triple-differential dijet measurement on the proton PDFs are demonstrated by including the cross section measurement in a PDF fit together with inclusive DIS cross sections from the HERA experiments. The PDF determination in this thesis follows the procedure applied in the original HERAPDF 2.0 publication [23]. Adjustments and improvements are applied where appropriate.

XFitter [24] is an open source framework to fit the PDFs to experimental data. The PDF evolution is based on the DGLAP [25–27] equations. To ensure consistency between the HERA DIS and the dijet cross section calculations, the fits are performed at NLO as the latter are only available at that order. The DIS cross sections are calculated by the QCDNUM software [95].

The PDF parametrization is based on the HERAPDF 2.0 approach [23]. After including the dijet data, a parametrization study is performed resulting in a slightly different parametrization which was found to better describe the included dijet data. The most notable difference is the neglection of the negative gluon term, a supplemental term allowing the gluon PDF to become negative at low x. No improvement of the fit result is found when including it. This is presumably due to the minimum Q^2 cut imposed on the DIS data, which has been adapted from $Q^2_{\text{min}} = 3.5\,\text{GeV}^2$ to $Q^2_{\text{min}} = 7.5\,\text{GeV}^2$ following a recent PDF study of inclusive jets [16]. However, this additional term in the parametrization of the gluon PDF is considered when deriving the parametrization uncertainties.

The parametrization of the PDFs is defined at the starting scale Q^2_0 which is set to $Q^2_0 = 1.9\,\text{GeV}^2$ and the five independent PDFs $xu_v(x)$, $xd_v(x)$, $xg(x)$, $x\bar{U}(x)$ and $x\bar{D}(x)$ are parametrized as follows:

$$
\begin{aligned}
xg(x) &= A_g x^{B_g}(1-x)^{C_g}(1+E_g x^2) \\
xu_v(x) &= A_{u_v} x^{B_{u_v}}(1-x)^{C_{u_v}}(1+D_{u_v} x) \\
xd_v(x) &= A_{d_v} x^{B_{d_v}}(1-x)^{C_{d_v}} \\
x\bar{U}(x) &= A_{\bar{U}} x^{B_{\bar{U}}}(1-x)^{C_{\bar{U}}}(1+D_{\bar{U}} x) \\
x\bar{D}(x) &= A_{\bar{D}} x^{B_{\bar{D}}}(1-x)^{C_{\bar{D}}}
\end{aligned}
$$

Actually, not all of those parameters are fitted. The normalization parameters A_g, A_{u_v} and A_{d_v} are calculated using the QCD sum rules. $B_{\bar{U}} = B_{\bar{D}}$ and $A_{\bar{U}} = A_{\bar{D}}(1-f_s)$ ensure the same normalization for the $\bar{u}$ and $\bar{d}$ PDF for the $x \to 0$ region. The strangeness fraction is set to $f_s = 0.40$. The generalized-mass variable-flavor-number-scheme as described in [18, 96] is used and the strong coupling constant is set to $\alpha_s(M_Z) = 0.1180$.

6.2.1 Treatment of Uncertainties in the PDF Fit

The uncertainty of the PDFs is subdivided into three independent sources, which are evaluated separately and finally added in quadrature to obtain the total uncertainty. This procedure was developed by HERAPDF [23] and is followed in this thesis.

Experimental Uncertainties They originate from statistical and systematic uncertainties of the data and are propagated to the PDFs using the Hessian eigenvector method [97]. The Hessian matrix is defined by the second derivatives of the fitted PDF parameters at the χ^2 minimum. The matrix is diagonalized and the eigenvectors are computed as depicted in Fig. 6.2.

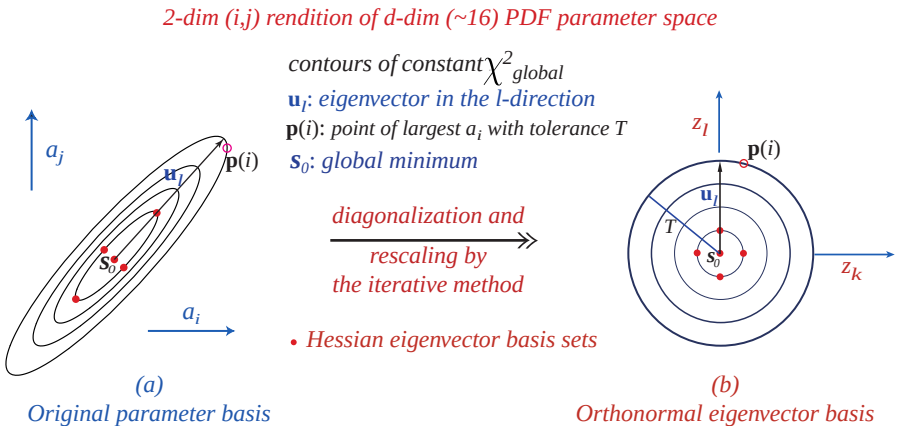

Figure 6.2: Transformation from the original parameter basis to the orthonormal eigenvector basis. The uncertainty of the PDFs can be propagated to a physical quantity X using these eigenvector PDF sets which are mutually uncorrelated. Taken from [97].

Using an iterative method, the downwards and upwards variation of each eigenvector corresponding to $\chi^2 = \chi^2_{\text{min}} + 1$ is calculated. Since the eigenvectors are orthogonal, the eigenvector variation PDFs correspond to independent sources of uncertainty on the PDFs. The asymmetric uncertainties ΔX^+_{exp} and ΔX^-_{exp} on a quantity X are evaluated as

$$\Delta X^+_{\text{exp}} = \sqrt{\sum_i^{N_{\text{EV}}} \left[\max(X_i^{\text{up}} - X_0, X_i^{\text{dn}} - X_0, 0)\right]^2}$$
$$\Delta X^-_{\text{exp}} = \sqrt{\sum_i^{N_{\text{EV}}} \left[\min(X_i^{\text{up}} - X_0, X_i^{\text{dn}} - X_0, 0)\right]^2},$$

where X_0 describes the central prediction and X_i^{up} and X_i^{dn} denote the result using the upwards and downwards variations of the eigenvector i of the N_{EV} eigenvectors in the PDF set. X may be a cross section calculation or the PDFs themselves.

Model Uncertainties The uncertainties of several input parameters in the PDF fits are combined into one PDF model uncertainty. For the evaluation of the model uncertainties, the following variations on the input parameters are considered, following the prescription of the HERAPDF 2.0 publication:

- The assumption is made that the shape of the strange quark PDF follows the shape of the down-like sea quark PDF. Thus, the strange quark PDF is not fitted but

determined as a fraction of the $x\overline{D}$ PDF. The strangeness fraction f_s, by default set to 0.40, is varied between 0.30 and 0.50.

- The b-quark mass is set to 4.5 GeV and varied between 4.25 GeV and 4.75 GeV.
- The c-quark mass, set by default to 1.47 GeV, is varied between 1.41 GeV and 1.53 GeV.
- The minimum Q^2 value for DIS data used in the fit, $Q^2_{\text{min}} = 7.5\,\text{GeV}^2$, is varied to $Q^2_{\text{min}} = 5.0\,\text{GeV}^2$ and $Q^2_{\text{min}} = 10.0\,\text{GeV}^2$.

The model uncertainty is evaluated in the same way as the PDF eigenvectors. The variation of each input parameter is treated as an independent variation leading to the definition of the asymmetric model uncertainty as:

$$\Delta X^{+}_{\text{mod}} = \sqrt{\sum_{i}^{N_{\text{par}}} \left[\max(X_i^{\text{up}} - X_0, X_i^{\text{dn}} - X_0, 0)\right]^2}$$
$$\Delta X^{-}_{\text{mod}} = \sqrt{\sum_{i}^{N_{\text{par}}} \left[\min(X_i^{\text{up}} - X_0, X_i^{\text{dn}} - X_0, 0)\right]^2}$$

Parametrization Uncertainty To estimate the influence of the parametrization choice on the fit result, a more flexible functional form is used. Using the general parametrizations for the gluon PDF

$$xg(x) = A_g x^{B_g}(1-x)^{C_g}(1 + D_g x + E_g x^2) - A'_g x^{B'_g}(1-x)^{C'_g}$$

and the quark PDFs

$$xf(x) = A_f x^{B_f}(1-x)^{C_f}(1 + D_f x + E_f x^2),$$

it is studied if the inclusion of additional parameters in the fit yields a different result. Each parameter is successively added in the PDF fit and the envelope of all changes to the PDF shape are combined into one parametrization uncertainty. Furthermore, the variation of the starting scale Q_0^2 to 1.6 GeV^2 and 2.2 GeV^2 is treated as a parametrization variation.

$$\Delta X^{+}_{\text{par}} = \max_{i}^{n} \left[X^i - X^0, 0\right]$$
$$\Delta X^{-}_{\text{par}} = \max_{i}^{n} \left[X^0 - X^i, 0\right]$$

6.2.2 Definition of the Goodness-of-Fit Estimator

In the PDF fit, a least-squares method is employed. The χ^2 is calculated with the data points D_i and the theoretical prediction T_i. The K correlated systematic uncertainties β_k are treated using nuisance parameters r_k, which are applied to the theoretical prediction in order to avoid the bias from multiplicative data uncertainties, see [98]. The χ^2 is defined as

$$\chi^2 = \sum_{ij}^{N} \left(D_i - T_i - \sum_{k}^{K} r_k \beta_{ik} \right) \mathrm{C}_{ij}^{-1} \left(D_j - T_j - \sum_{k}^{K} r_k \beta_{jk} \right) + \sum_{k}^{K} r_k^2 + \sum_{j} \ln \frac{\Delta_{i,\mathrm{stat}}^2 D_j T_j + \Delta_{i,\mathrm{uncor}}^2 T_i^2}{\left(\Delta_{i,\mathrm{stat}}^2 + \Delta_{i,\mathrm{uncor}}^2 \right) D_i^2}$$

where C_{ij}^{-1} is the inverse covariance matrix of the uncorrelated or partly correlated systematic and statistical uncertainties and Δ_{stat} and Δ_{uncor} are the diagonal entries of the statistical and uncorrelated covariance matrix. All fully correlated uncertainties are treated using nuisance parameters. The logarithmic term in the χ^2 definition is an additional correction which arises from the transition of the Gaussian distribution to the χ^2 formula, if the uncertainties are not constant.

More information can be found in [23, 24]. An advantage of such a χ^2 definition with nuisance parameters is the possibility to study the pulls of each systematic source after the fit.

6.2.3 Treatment of Systematic Uncertainties

Since the correlated sources of uncertainty are treated using nuisance parameters in the χ^2 formula, the pull of each source can be examined. Tab. 6.1 shows the 27 sources of systematic uncertainty and their pulls. Most of the systematic sources shift by less than one standard deviation. One source exhibits a larger shift. While this is not surprising because of the Gaussian distribution of the pulls, no unambiguous reason for this behavior could be identified. However, the size of this uncertainty source is small compared to the dominant sources of uncertainty and no significant influence on the fit is expected.

Table 6.1: Nuisance parameters obtained after the PDF fit. The shift of each nuisance parameter is reported in standard deviations of its represented uncertainty source. Most of the pulls cause shifts by less than a standard deviation.

Systematic source	Shift in σ	Systematic source	Shift in σ
AbsoluteScale	-0.33	RelativeFSR	0.97
AbsoluteStat	-0.29	RelativeStatEC2	0.52
AbsoluteMPFBias	-0.57	RelativeStatHF	0.08
Fragmentation	-0.22	RelativeStatFSR	-0.23
SinglePionECAL	-0.20	PileUpDataMC	0.03
SinglePionHCAL	-0.75	PileUpPtRef	-0.27
FlavorQCD	-0.77	PileUpPtBB	-2.57
RelativeJEREC1	0.12	PileUpPtEC1	1.11
RelativeJEREC2	0.04	PileUpPtEC2	0.37
RelativeJERHF	0.14	PileUpPtHF	0.00
RelativePtBB	-0.74	NPErr	1.04
RelativePtEC1	0.54	JERerr	0.62
RelativePtEC2	-0.72	Lumi	0.84
RelativePtHF	0.00		

6.3 PDFs Constraints of the Triple-Differential Dijet Cross Section

The quality of the fit with and without including the dijet measurement is reported in Table 6.2. The partial χ^2 per data point for each data set as well as the χ^2 per n_{dof} for all data sets demonstrate the compatibility of the CMS dijet measurement and the DIS data from the HERA experiments.

The resulting PDFs for the gluon, u valence, d valence and sea quark for a fit with and without the CMS dijet data are shown next to each other in Fig. 6.3 and Fig. 6.4 with a breakdown of the three sources of uncertainty. A direct comparison of the PDFs with total uncertainties is presented in Fig. 6.5. The uncertainty breakdown in Fig. 6.3 reveals the large impact of the dijet data: The uncertainty of the gluon PDF is reduced over almost the whole range in x. In the high-x region, the experimental uncertainty is reduced, while most changes in the low-x region can be attributed to the model uncertainties. The large changes of uncertainties come along with a noticeable change of the gluon PDF shape. Compared to the fit with HERA DIS data alone, the gluon PDF decreases at medium x and increases at high x. Similar changes were observed before, e. g. in [16].

The changes of the sea quark PDF, defined as $x\Sigma = 2(x\overline{U} + x\overline{D} + x\overline{s})$, are much less pronounced. The experimental and model uncertainties are reduced in the high-x region. The u valence and d valence quark PDFs also show uncertainty reductions over a large

Table 6.2: The partial χ^2 for each data set in the HERA DIS (middle section) or the combined fit including the triple-differential dijet data (right section) are shown. The bottom two lines show the total χ^2 and χ^2/n_{dof}. The difference between the sum of all χ^2_{p} and the total χ^2 for the combined fit is attributed to the nuisance parameters which are reported separately.

		HERA data		**HERA & CMS data**	
data set	n_{data}	χ^2_{p}	$\chi^2_{\text{p}}/n_{\text{data}}$	χ^2_{p}	$\chi^2_{\text{p}}/n_{\text{data}}$
CC HERA-I+II e^-p	42	54.64	1.30	51.88	1.24
CC HERA-I+II e^+p	39	37.47	0.96	37.47	0.96
NC HERA-I+II e^-p	159	216.72	1.36	219.66	1.38
NC HERA-I+II e^+p $E_{\text{p}} = 460\,\text{GeV}$	187	204.97	1.10	205.65	1.10
NC HERA-I+II e^+p $E_{\text{p}} = 575\,\text{GeV}$	234	196.90	0.84	198.39	0.85
NC HERA-I+II e^+p $E_{\text{p}} = 820\,\text{GeV}$	63	61.68	0.98	61.83	0.98
NC HERA-I+II e^+p $E_{\text{p}} = 920\,\text{GeV}$	332	376.03	1.13	401.72	1.21
CMS Triple-Differential Dijets	111	—	—	99.10	0.89
data set(s)	n_{dof}	χ^2	χ^2/n_{dof}	χ^2	χ^2/n_{dof}
HERA data	1043	1208.10	1.16	—	—
HERA & CMS data	1154	—	—	1341.54	1.16

x range, most pronounced at high x. The decrease of the experimental uncertainty is sizable for high x, while the main differences result from the parametrization uncertainty, which is significant for the fit with HERA DIS data alone, but almost completely vanishes in the fit including the CMS dijet data.

Figure A.14 in the appendix shows the PDFs after the evolution to the scale $Q^2 = 10\,000\,\text{GeV}$, which is close to the scale of the measurement. The PDFs exhibit the same features compared to the starting scale Q_0, but the uncertainty reduction is even more pronounced. Finally, an overview of the gluon, sea, u valence and d valence quark PDFs is given in Fig. 6.6.

The large impact of the dijet data on the PDFs is partly caused by the limited precision of the fit using HERA DIS data alone. In a global PDF fit, constraints from other measurements already improve the PDFs and a less strong impact is expected. However, it was shown in Sec. 5.9 that the data are sufficiently precise to discriminate between the predictions using different global PDF sets. Thus, significant improvements of the global PDF sets are expected likewise.

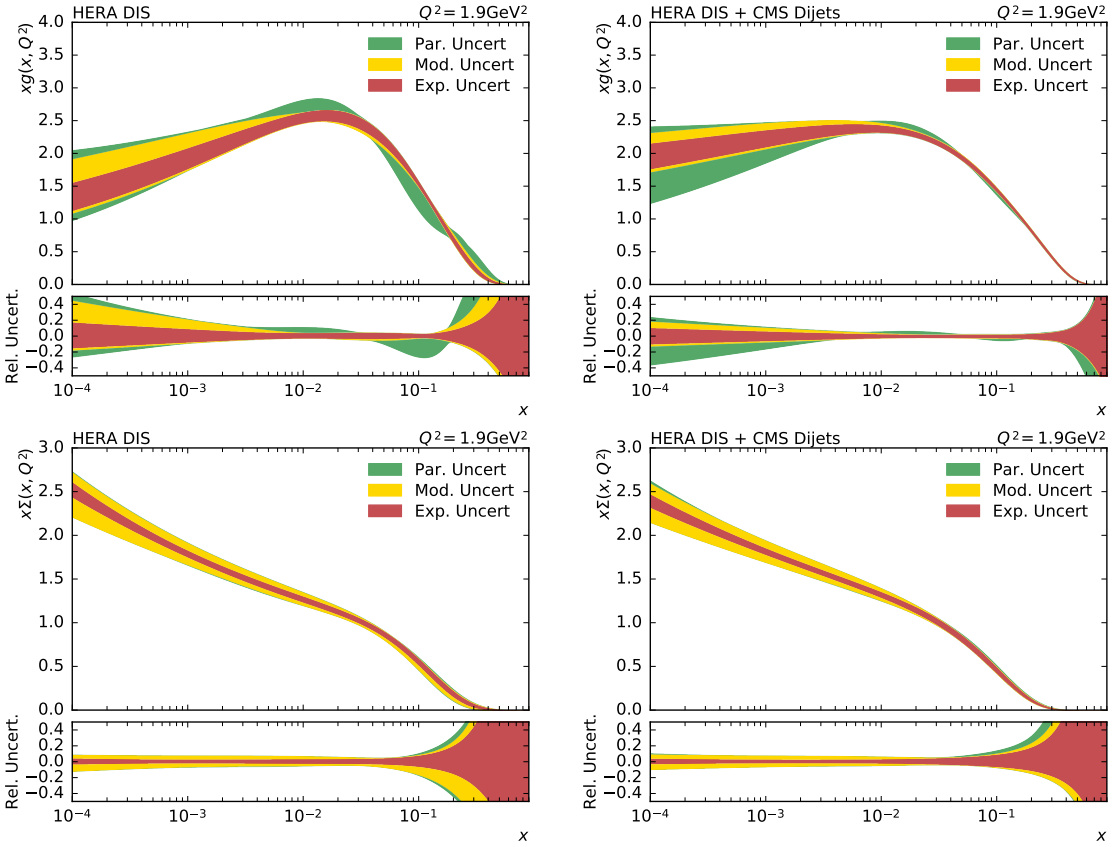

Figure 6.3: The gluon (top) and sea quark (bottom) PDFs as a function of x as derived from HERA inclusive DIS data alone (left) and in combination with the CMS dijet data (right). The PDFs are shown at the starting scale $Q^2 = 1.9\,\mathrm{GeV}^2$. The experimental (inner band), model (middle band) and parametrization uncertainties (outer band) are added quadratically to give the total uncertainty.

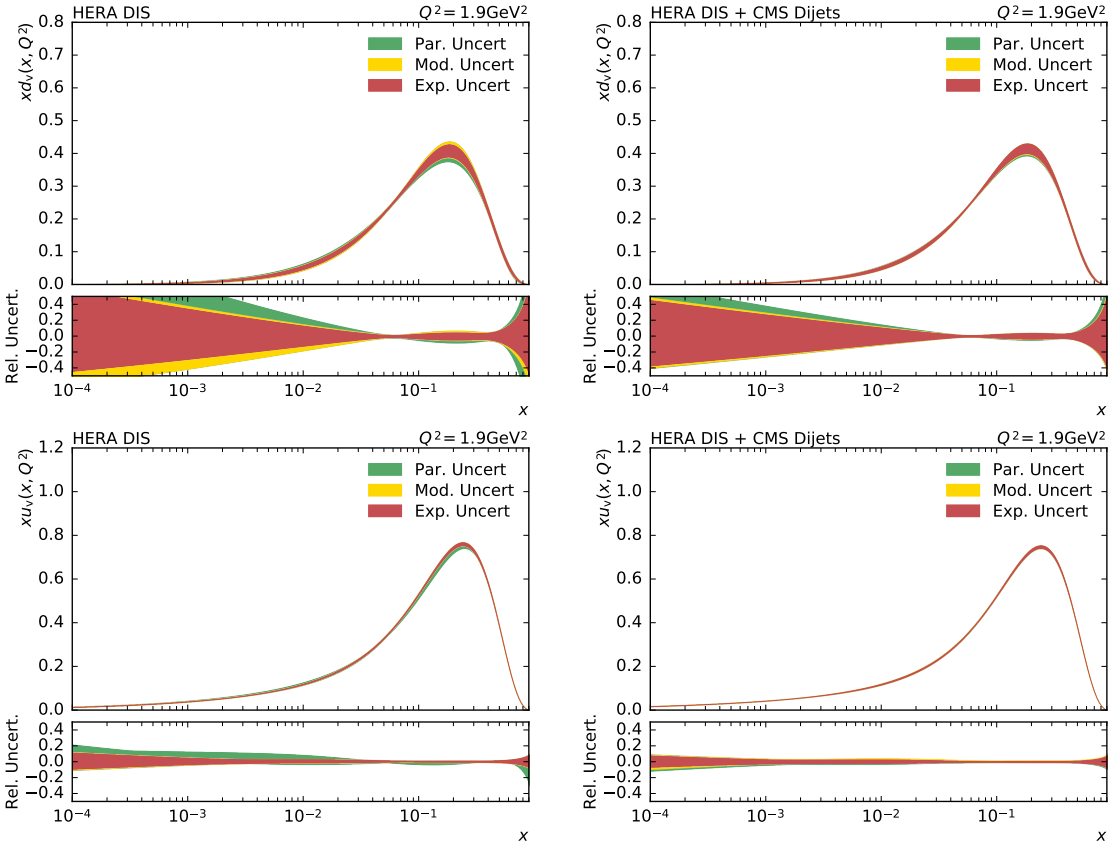

Figure 6.4: The d valence quark (top) and u valence quark (bottom) PDFs as a function of x as derived from HERA inclusive DIS data alone (left) and in combination with the CMS dijet data (right). The PDFs are shown at the starting scale $Q^2 = 1.9\,\text{GeV}^2$. The experimental (inner band), model (middle band) and parametrization uncertainties (outer band) are added quadratically to give the total uncertainty.

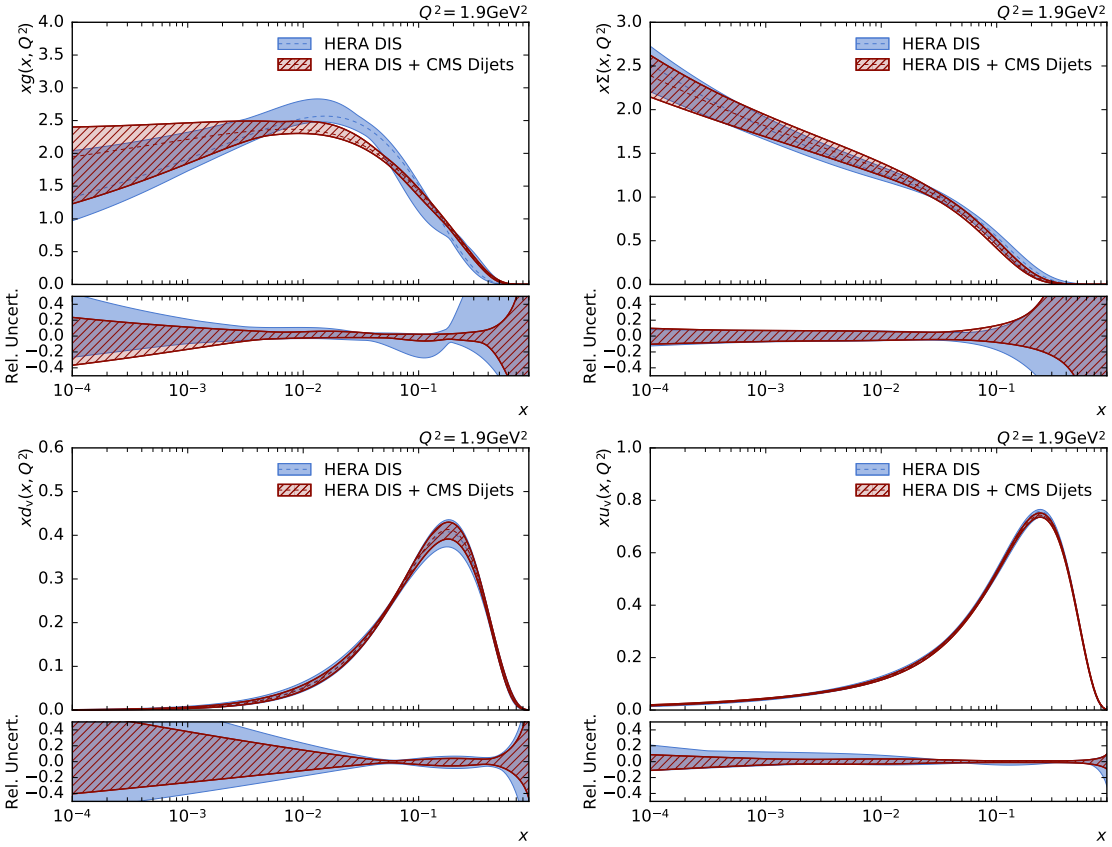

Figure 6.5: The gluon (top left), sea quark (top right), d valence quark (bottom left) and u valence quark (bottom right) PDFs as a function of x as derived from HERA inclusive DIS data alone (hatched band) and in combination with CMS dijet data (solid band). The PDFs are shown at the starting scale $Q^2 = 1.9\,\mathrm{GeV}^2$. The total uncertainty of the PDFs is shown.

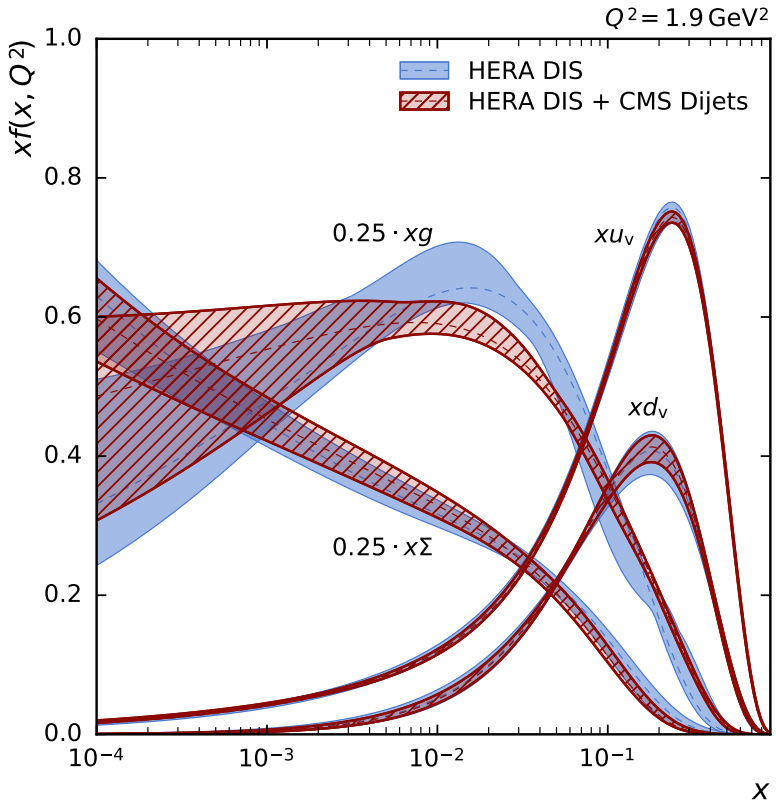

Figure 6.6: Overview of the gluon, sea, u valence and d valence quark PDFs before (hatched band) and after (solid band) including the CMS dijet data in the fit. The PDFs are shown at the starting scale $Q^2 = 1.9\,\mathrm{GeV}^2$ with total uncertainties. The uncertainties of the PDFs, especially those of the gluon PDF, are significantly reduced. Furthermore, a change of the gluon PDF shape is observed, resulting in a larger gluon PDF at high x after including the dijet data.

6.4 Simultaneous Fit of PDFs and Strong Coupling Constant

The measurement of triple-differential dijet cross sections does not only provide constraints on the PDFs, but also on the strong coupling constant. As shown in the preceding part of this chapter, the region containing boosted dijet events is the most sensitive regarding the PDFs. However, the α_s sensitivity is higher in the central bin, in which highest transverse momenta are reached, and the experimental uncertainties are smallest.

Due to the known correlation between the different phase space regions in this measurement, this can be exploited by performing a simultaneous fit of the PDFs and the strong coupling. Since the HERA DIS data are less sensitive to the $\alpha_s(M_Z)$ value, a consistent value of $\alpha_s(M_Z)$ is extracted from the CMS dijet data as no other data enter the fit. This is a major advantage of the applied method. A disadvantage of such a fit is that, unlike in a global PDF fit, further constraints from other measurements on the PDFs are neglected.

The fit is set up in the same manner as in the previous PDF studies. One additional free parameter is included: the strong coupling constant $\alpha_s(M_Z)$. The obtained value for the strong coupling constant reads

$$\alpha_s(M_Z) = 0.1194^{+0.0015}_{-0.0015}(\text{exp})^{+0.0002}_{-0.0002}(\text{mod})^{+0.0002}_{-0.0004}(\text{par}),$$

where the experimental uncertainty accounts for all sources of uncertainties of the HERA and CMS data sets. NP uncertainties, which were also included in the fit, are accounted for in the experimental uncertainty. However, the NP uncertainties are comparably small and do not have a significant influence on the total experimental uncertainty when added quadratically. Furthermore, model and parametrization uncertainties were calculated in the same way as in the PDF determination.

The consideration of scale uncertainties in a PDF fit is an open issue in the PDF community. Therefore, two different methods to evaluate the scale uncertainty on $\alpha_s(M_Z)$ were studied. Similar to what is reported in Sec. 4.4.1, the renormalization and factorization scales were varied in the calculation of the dijet data. The simultaneous fit was repeated for each variation. The envelope of the best fit $\alpha_s(M_Z)$ values for all variations is taken as uncertainty: $\Delta\alpha_s(M_Z) = ^{+0.0026}_{-0.0016}$ (scale).

The second procedure is analogous to the method which was applied in previous determinations of $\alpha_s(M_Z)$ e. g. in [16, 17]. The PDFs are derived for a series of fixed values of $\alpha_s(M_Z)$. Using this series, the best fit $\alpha_s(M_Z)$ value of the dijet data is determined for each scale variation. As before, the envelope of all variations is taken as scale uncertainty: $\Delta\alpha_s(M_Z) = ^{+0.0031}_{-0.0019}$ (scale). Since this uncertainty is the most consistent to compare with previous determinations of $\alpha_s(M_Z)$, it is reported as the default one.

The determined value of $\alpha_s(M_Z)$ is in good agreement with the world average of $\alpha_s(M_Z) = 0.1181 \pm 0.0013$ determined by the PDG [2]. The dominant source of uncertainty is due to the scale variations. When the NNLO dijet calculations will become available, a major improvement is expected due to the reduced scale dependence.

CHAPTER 7

Conclusion

The PDFs of the proton enter almost all cross section calculations at the LHC. Thus, a profound knowledge of the PDFs is crucial for all precision measurements. However, in particular the high-x region of the PDFs is not yet well known and exhibits large uncertainties.

This thesis presents a new dijet analysis, specifically developed to provide constraints on the PDFs in the best possible way. For the first time, triple-differential dijet cross sections are measured at the LHC. Constraints on the PDFs are derived and the strong coupling constant is determined.

The measurement has been performed with the CMS detector at a center-of-mass energy of 8 TeV using the complete data set recorded in 2012. The cross sections are measured differentially as a function of the average transverse momentum, the rapidity separation, and the boost of the dijet pair. Trigger, reconstruction and detector effects were thoroughly studied. The measured cross sections have been corrected for detector effects in an iterative unfolding procedure. The unfolded cross sections are compared with pQCD predictions at NLO accuracy which were corrected for non-perturbative effects. The data are well described by the predictions over many orders of magnitude, see Fig. 7.1 left. In phase space regions with boosted dijet events, in which the highest x of the PDFs are probed, the precise measurement discriminates between predictions of different global PDF sets and constraints on the PDFs can thus be provided.

The impact of the data on the PDFs is demonstrated by performing a PDF fit to DIS cross sections obtained from the HERA experiments and the dijet cross sections measured in this thesis. If the dijet data are considered, a harder gluon PDF is obtained and the overall uncertainties of the PDFs, especially those of the gluon PDF, are significantly reduced, see Fig. 7.1 right.

The strong coupling constant $\alpha_s(M_Z)$ has been determined together with the PDFs in a simultaneous fit. The obtained value reads

$$\alpha_s(M_Z) = 0.1194^{+0.0015}_{-0.0015}\,(\text{exp})\,^{+0.0002}_{-0.0002}\,(\text{mod})\,^{+0.0002}_{-0.0004}\,(\text{par})\,^{+0.0031}_{-0.0019}\,(\text{scale})$$

and is in agreement with the world average value of $\alpha_s(M_Z) = 0.1181 \pm 0.0013$ determined by the PDG [2]. The dominant uncertainty is of theoretical origin.

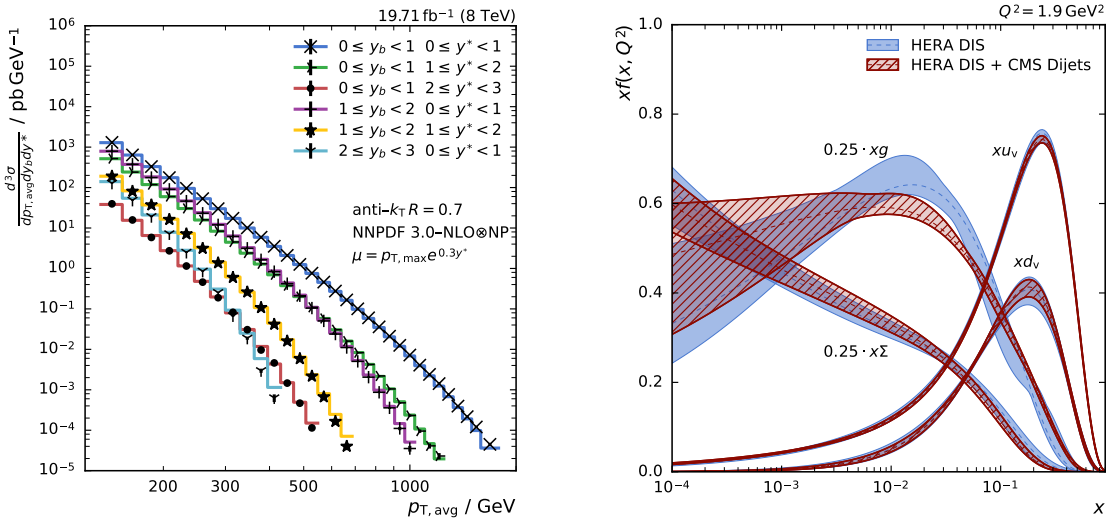

Figure 7.1: Left: The triple-differential dijet cross sections. The data are indicated by black markers, the NLO theory prediction by colored lines. Right: Overview of fitted PDFs with and without including the triple-differential dijet measurement.

The pioneering studies presented in this thesis prove that the triple-differential measurement of dijet cross sections using the chosen observables are an excellent approach to perform QCD precision studies.

A few areas provide opportunities for further improvement: When NNLO corrections for dijet calculations will become available in the near future, the accuracy of the cross section predictions will improve. Especially the $\alpha_s(M_Z)$ value obtained at NNLO accuracy will profit from reduced scale uncertainties. Moreover, it would allow to consider this measurement in the global $\alpha_s(M_Z)$ combination of the PDG.

Electroweak corrections become relevant at transverse momenta beyond 1 TeV. By considering them, the bins with highest $p_{T,avg}$ could be included in the fits as well.

The restart of the LHC at an increased center-of-mass energy of 13 TeV, coupled with a higher instantaneous luminosity, opens up a larger accessible phase space. When CMS has collected a sufficiently large data sample, it will be possible to extend this measurement to additional phase space regions and improve the overall precision.

APPENDIX A

Appendix

A.1 Subprocess Fraction in Dijet Production

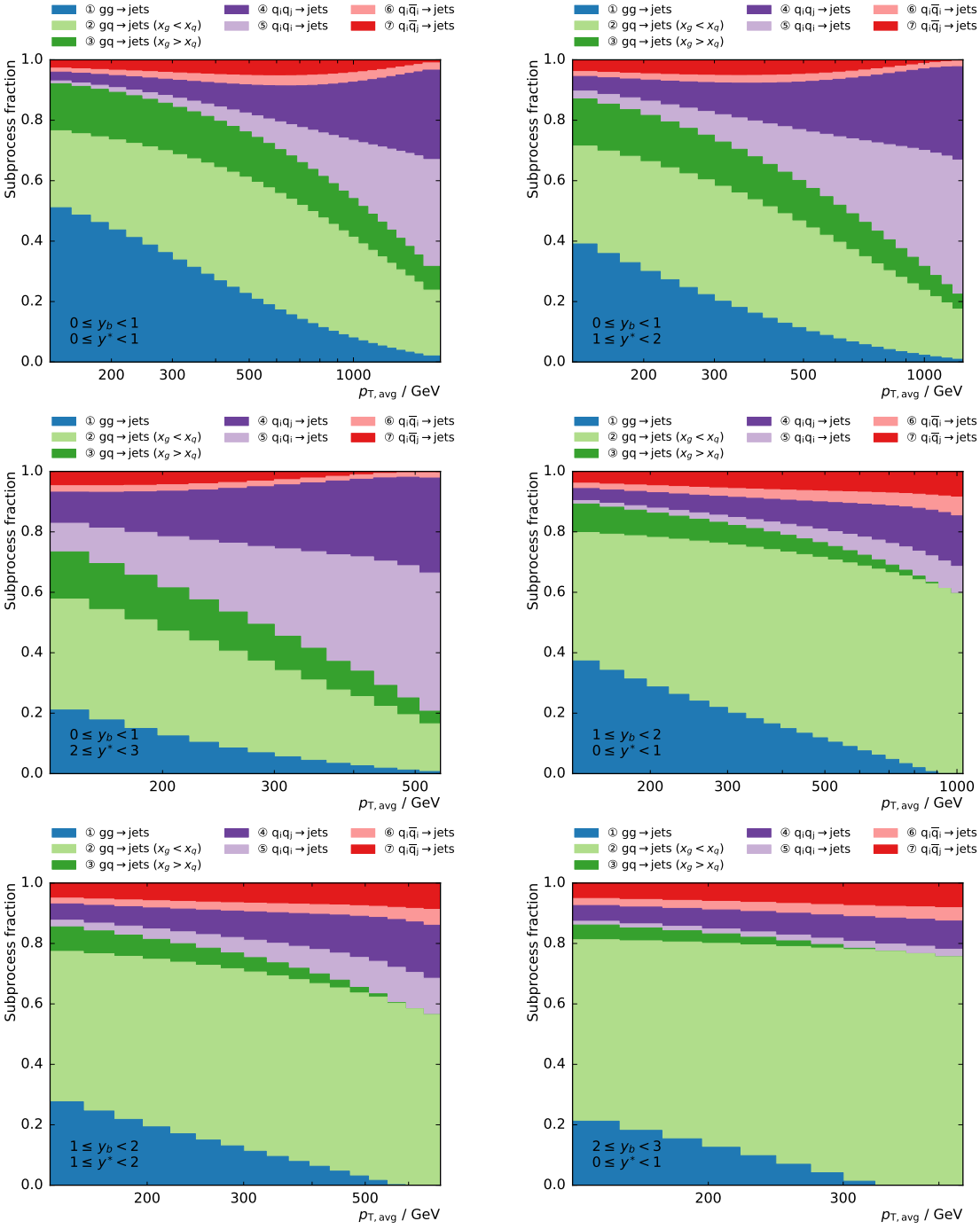

Figure A.1: The relevant subprocesses for dijet production can be reduced to seven subprocesses. There relative contribution to the total cross section in the six studied phase space regions is shown.

A.2 Monte Carlo Data Sets

The Monte Carlo studies in this analysis have beeen performed with the following samples:

Table A.1: The official MC production samples used in the analysis were generated in phase space slices in p_T and HT with the generators Pythia 8 and Madgraph, respectively. Madgraph was interfaced to Pythia 6 for the parton shower and hadronization of the events.

Generator	Identifier	Events	Cross Section pb
Pythia 8	/QCD_Pt-30to50_Tune4C_8TeV_pythia8/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	1 000 080	7.500×10^7
	/QCD_Pt-50to80_Tune4C_8TeV_pythia8/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	1 000 026	9.264×10^6
	/QCD_Pt-80to120_Tune4C_8TeV_pythia8/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	1 000 054	1.165×10^6
	/QCD_Pt-120to170_Tune4C_8TeV_pythia8/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	800 064	1.750×10^5
	/QCD_Pt-170to300_Tune4C_8TeV_pythia8/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	800 046	3.797×10^4
	/QCD_Pt-300to470_Tune4C_8TeV_pythia8/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	500 038	1.939×10^3
	/QCD_Pt-470to600_Tune4C_8TeV_pythia8/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	500 051	1.249×10^2
	/QCD_Pt-600to800_Tune4C_8TeV_pythia8/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	492 988	2.955×10^1
	/QCD_Pt-800to1000_Tune4C_8TeV_pythia8/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	400 059	3.87
	/QCD_Pt-1000to1400_Tune4C_8TeV_pythia8/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	400 050	0.803
	/QCD_Pt-1400to1800_Tune4C_8TeV_pythia8/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	200 070	0.363×10^{-1}
	/QCD_Pt-1800toInf_Tune4C_8TeV_pythia8/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	200 013	0.198×10^{-2}
Madgraph + Pythia 6	/QCD_HT-100To250_TuneZ2star_8TeV-madgraph-pythia/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	50 129 518	1.036×10^7
	/QCD_HT-250To500_TuneZ2star_8TeV-madgraph-pythia/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	27 062 078	2.760×10^5
	/QCD_HT-500To1000_TuneZ2star_8TeV-madgraph-pythia/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	30 599 292	8.426×10^3
	/QCD_HT-1000ToInf_TuneZ2star_8TeV-madgraph-pythia/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	13 843 863	2.040×10^2

A.3 All Sources of Jet Energy Correction Uncertainties

The following figures show the individual components of the jet energy correction uncertainties.

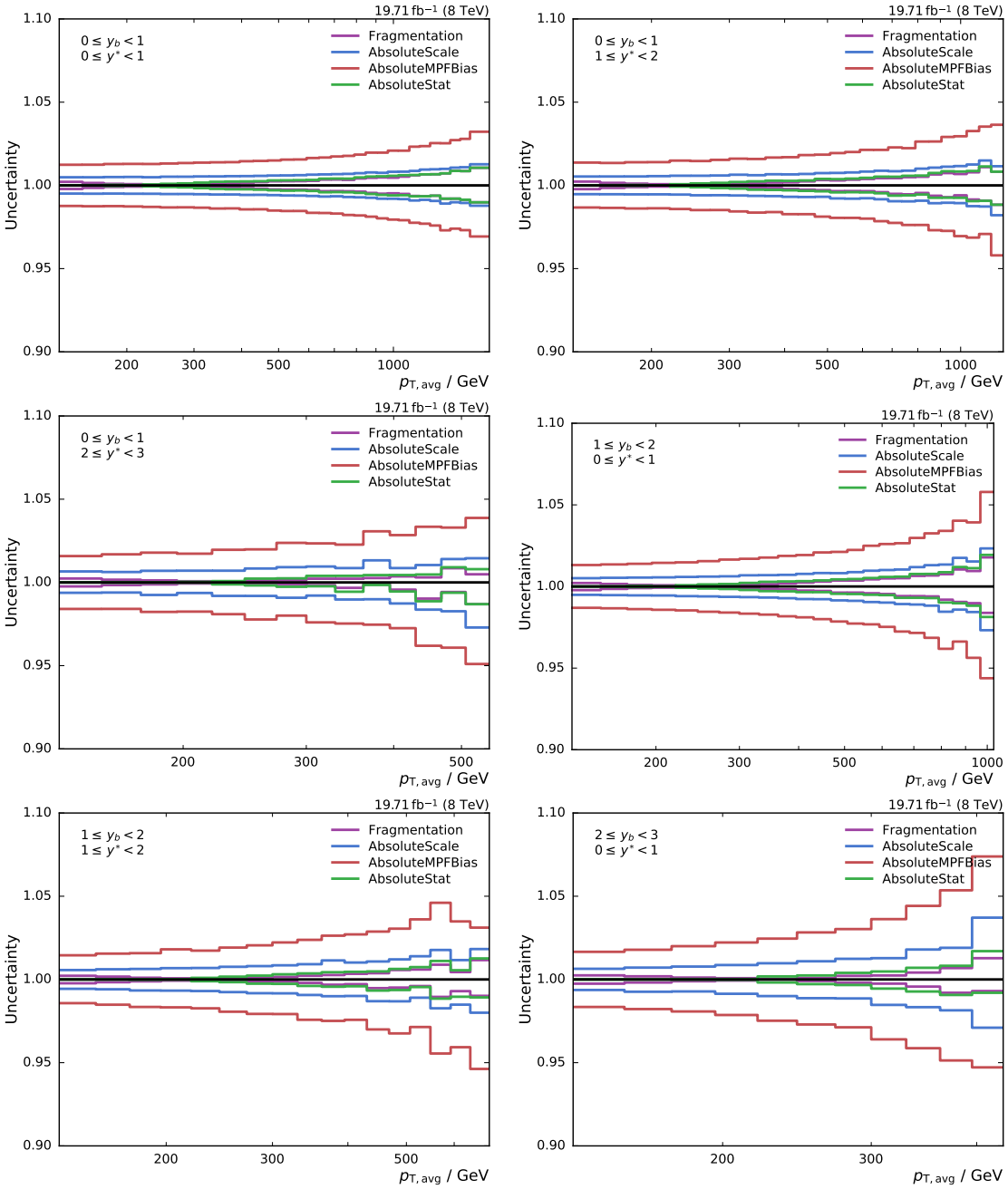

Figure A.2: The relative size of the jet energy scale uncertainties for the sources Fragmentation, AbsoluteScale, AbsoluteMPFBias, and AbsoluteStat are shown for all y^* and y_b bins.

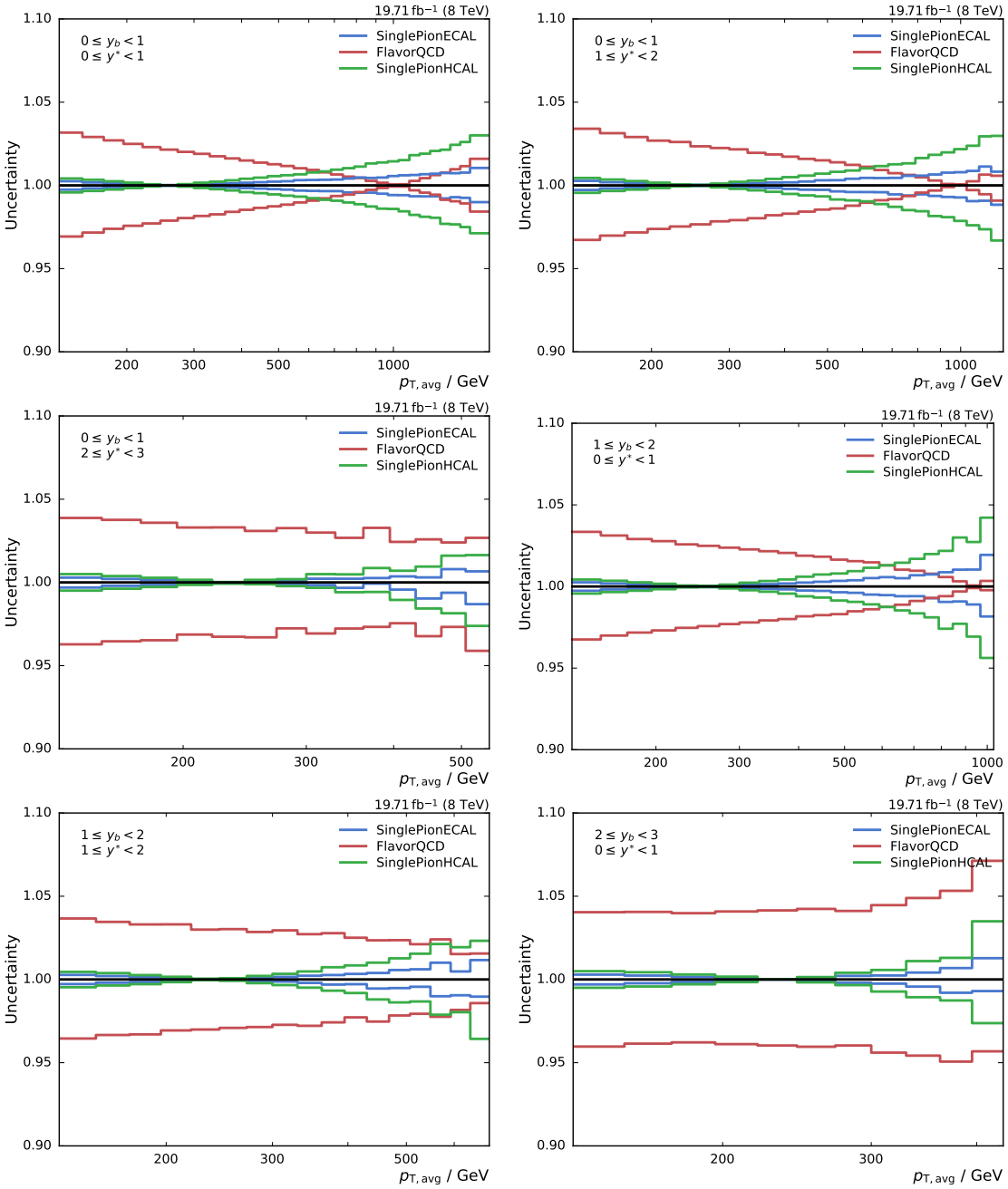

Figure A.3: The relative size of the jet energy scale uncertainties for the sources SinglePionECAL, SinglePionHCAL, and FlavorQCD are shown for all y^* and y_b bins.

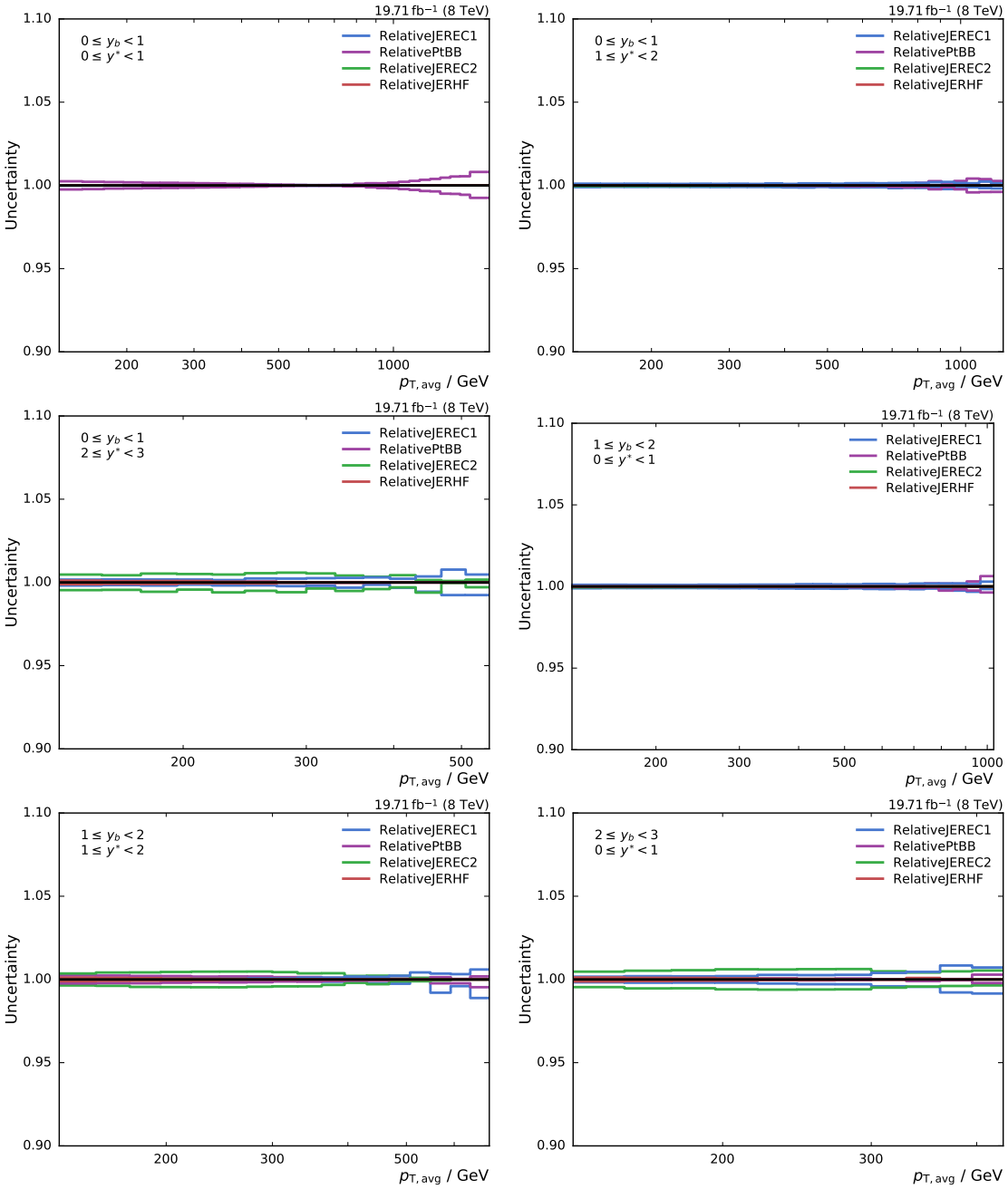

Figure A.4: The relative size of the jet energy scale uncertainties for the sources RelativeJEREC1, RelativeJEREC2, RelativePtBB, and RelativeJERHF are shown for all y^* and y_b bins.

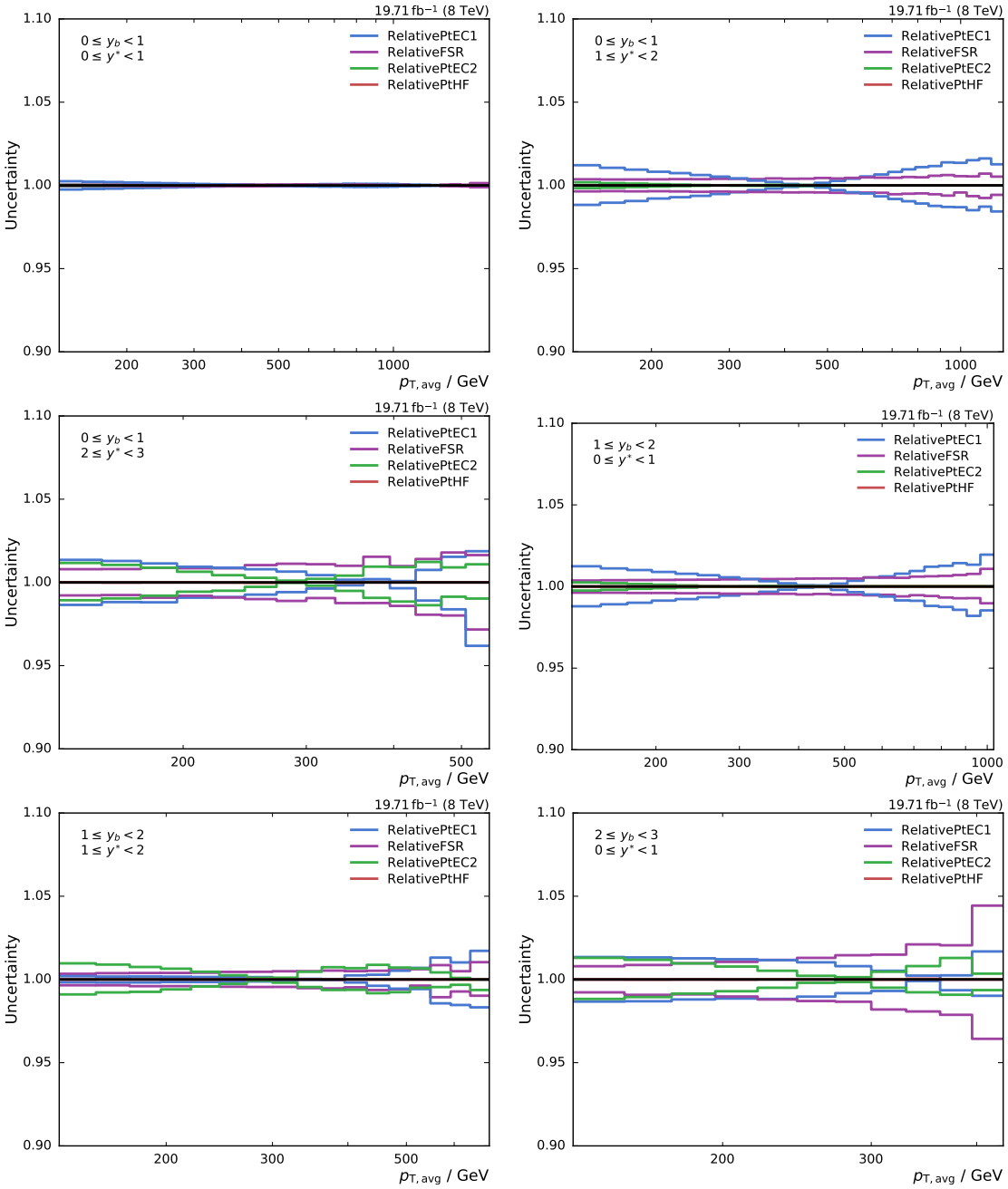

Figure A.5: The relative size of the jet energy scale uncertainties for the sources RelativePtEC1, RelativePtEC2, RelativeFSR, and RelativePtHF are shown for all y^* and y_b bins.

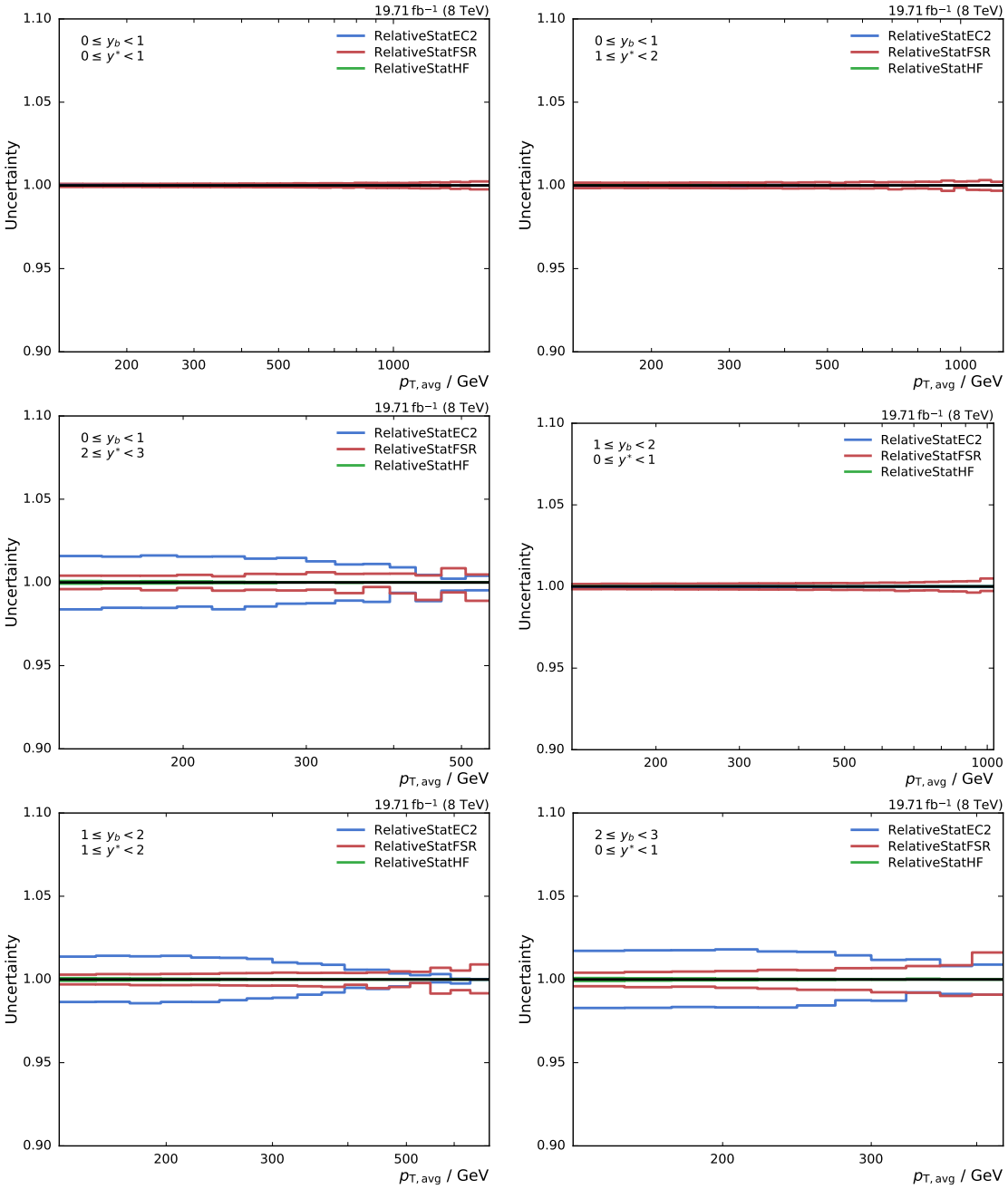

Figure A.6: The relative size of the jet energy scale uncertainties for the sources RelativeStatEC2, RelativeStatFSR, and RelativeStatHF are shown for all y^* and y_b bins.

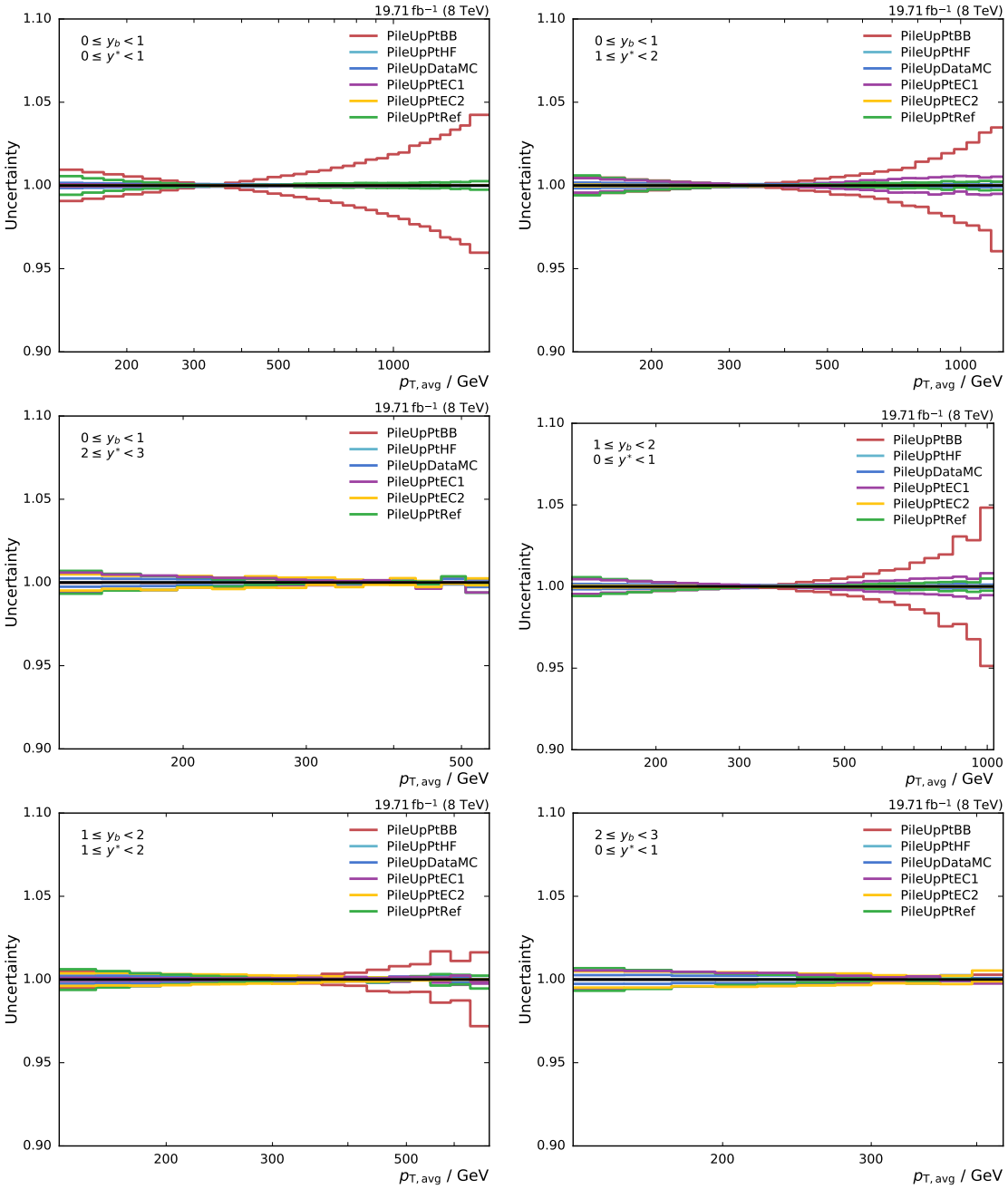

Figure A.7: The relative size of the jet energy scale uncertainties for the sources PileUpPtBB, PileUpPtHF, PileUpDataMC, PileUpPtEC1, PileUpPtEC2, and PileUpPtRef are shown for all y^* and y_b bins.

A.4 Additional Data and Theory Comparisons

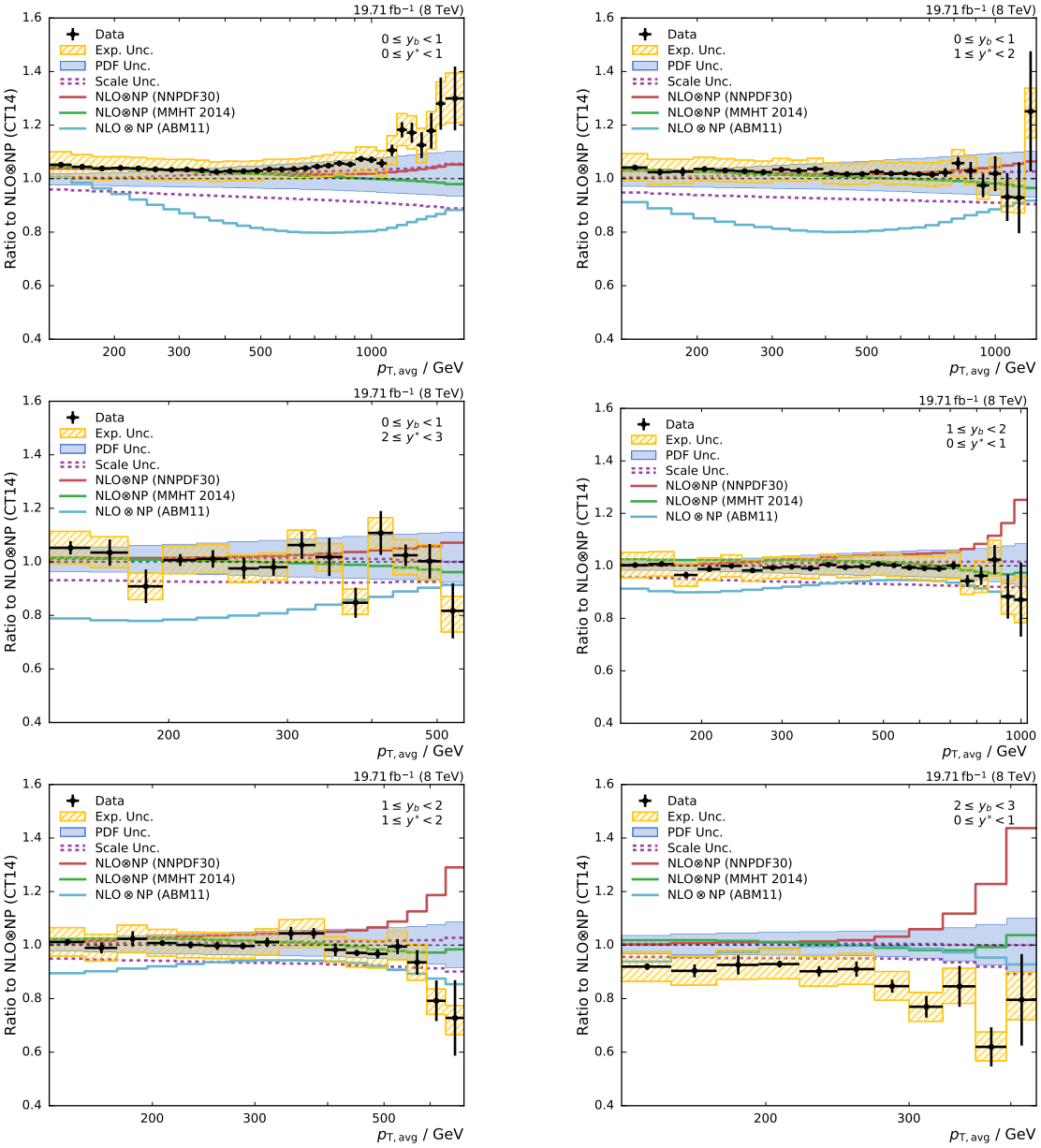

Figure A.8: Ratio of the triple-differential dijet cross sections to the theoretical prediction using the central value of the CT14 NLO PDF set for each bin in y^* and y_b respectively. The data points including statistical uncertainty are indicated by markers, the total experimental uncertainty is represented by the hatched band. The solid blue band indicates the PDF uncertainty and the continous colored lines the predictions of the cross sections calculated with other PDF sets.

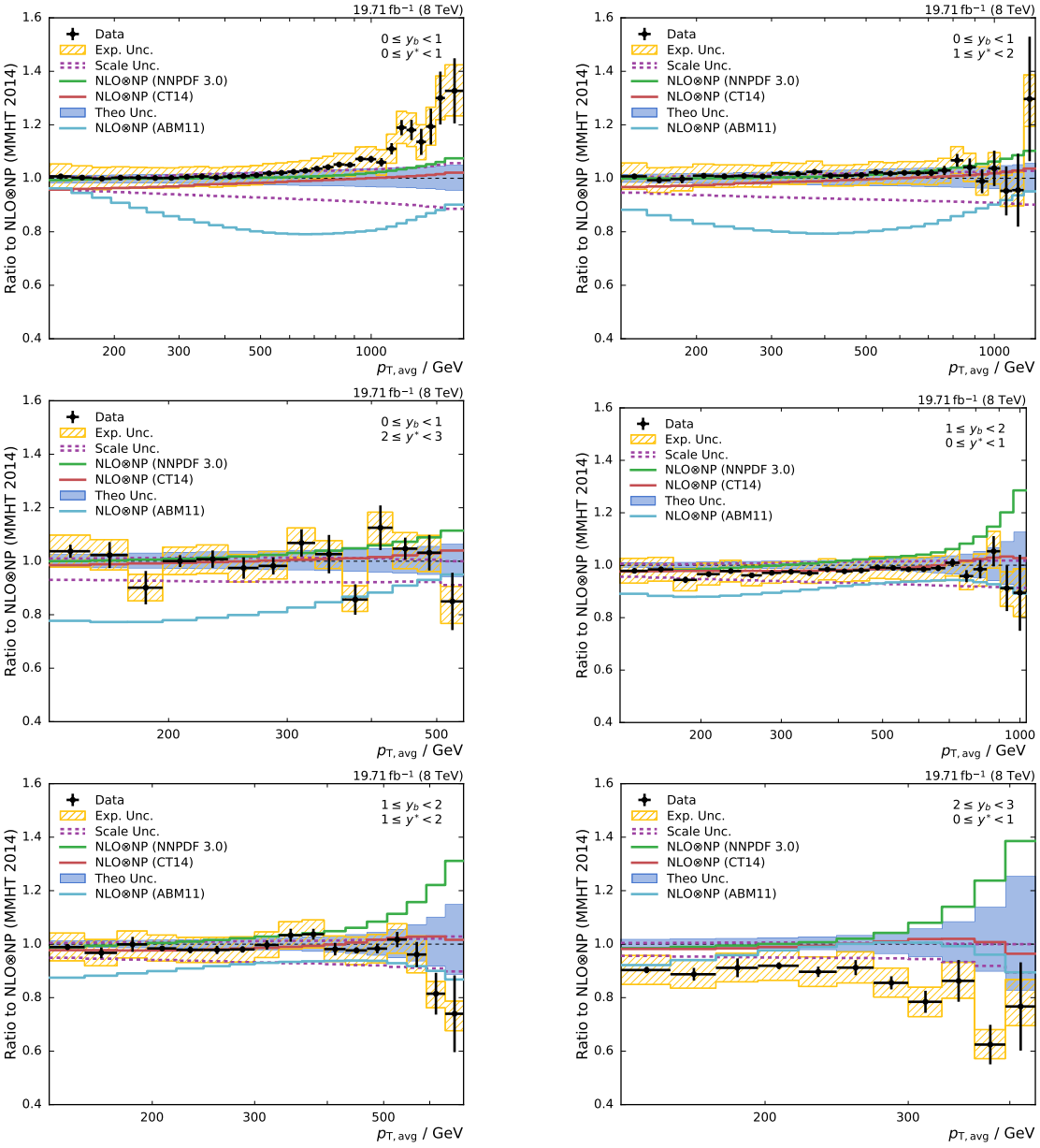

Figure A.9: Ratio of the triple-differential dijet cross sections to the theoretical prediction using the central value of the MMHT 2014 NLO PDF set for each bin in y^* and y_b respectively. The data points including statistical uncertainty are indicated by markers, the total experimental uncertainty is represented by the hatched band. The solid blue band indicates the PDF uncertainty and the continous colored lines the predictions of the cross sections calculated with other PDF sets.

A.5 PDF Correlations

Additional material for the correlation studies of Sec. 6.1.

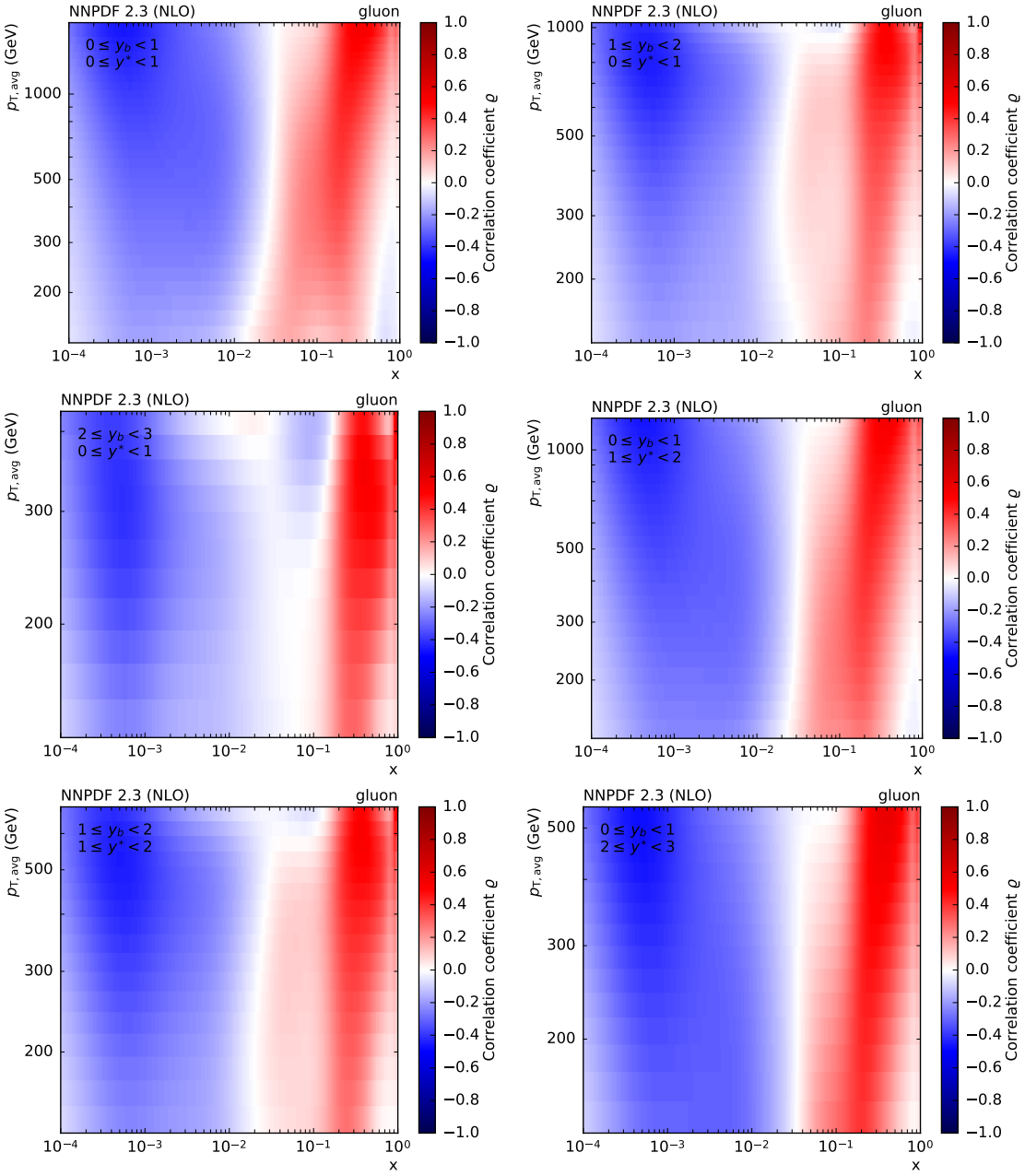

Figure A.10: The correlation coefficient between the triple-differential dijet cross section and the gluon PDF, as a function of the momentum fraction x of the proton and the energy scale μ of the hard process. The correlation is shown for the six bins in y^* and y_b of the measurement.

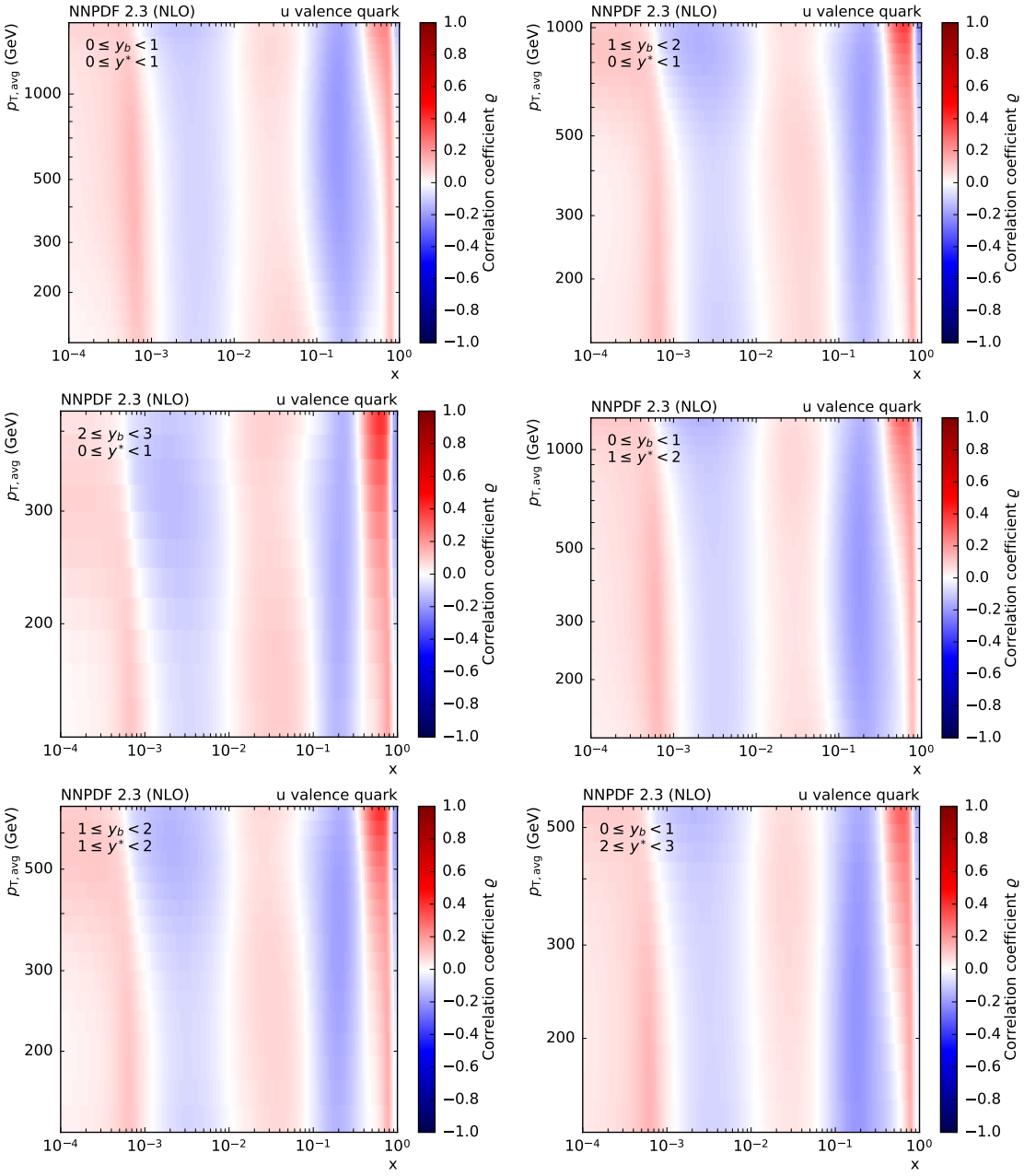

Figure A.11: The correlation coefficient between the triple-differential dijet cross section and the u valence quark PDF, as a function of the momentum fraction x of the proton and the energy scale μ of the hard process. The correlation is shown for the six bins in y^* and y_b of the measurement.

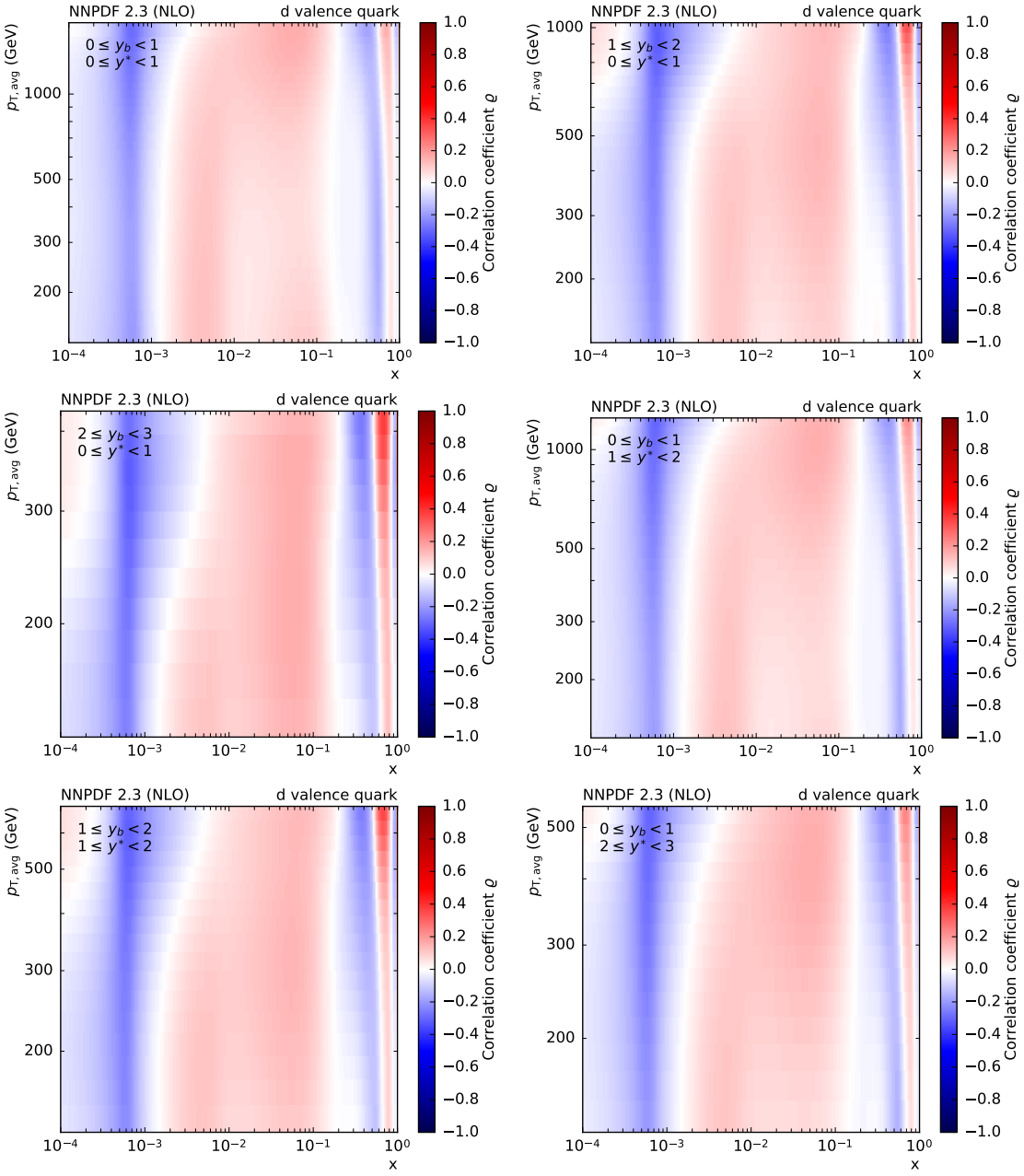

Figure A.12: The correlation coefficient between the triple-differential dijet cross section and the d valence quark PDF, as a function of the momentum fraction x of the proton and the energy scale μ of the hard process. The correlation is shown for the six bins in y^* and y_b of the measurement.

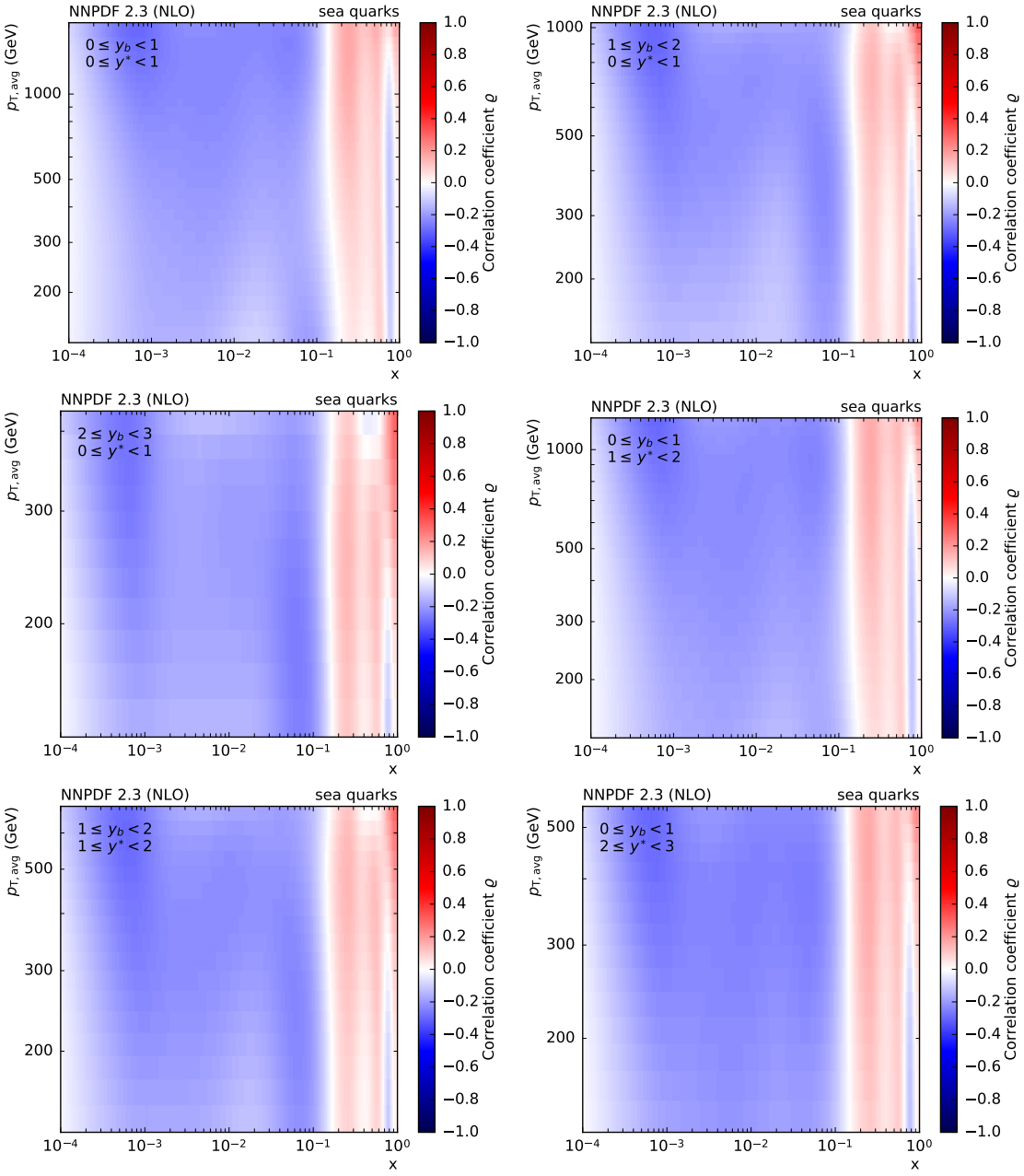

Figure A.13: The correlation coefficient between the triple-differential dijet cross section and the sea quarks PDF, as a function of the momentum fraction x of the proton and the energy scale μ of the hard process. The correlation is shown for the six bins in y^* and y_b of the measurement.

A.6 Additional PDF comparisons

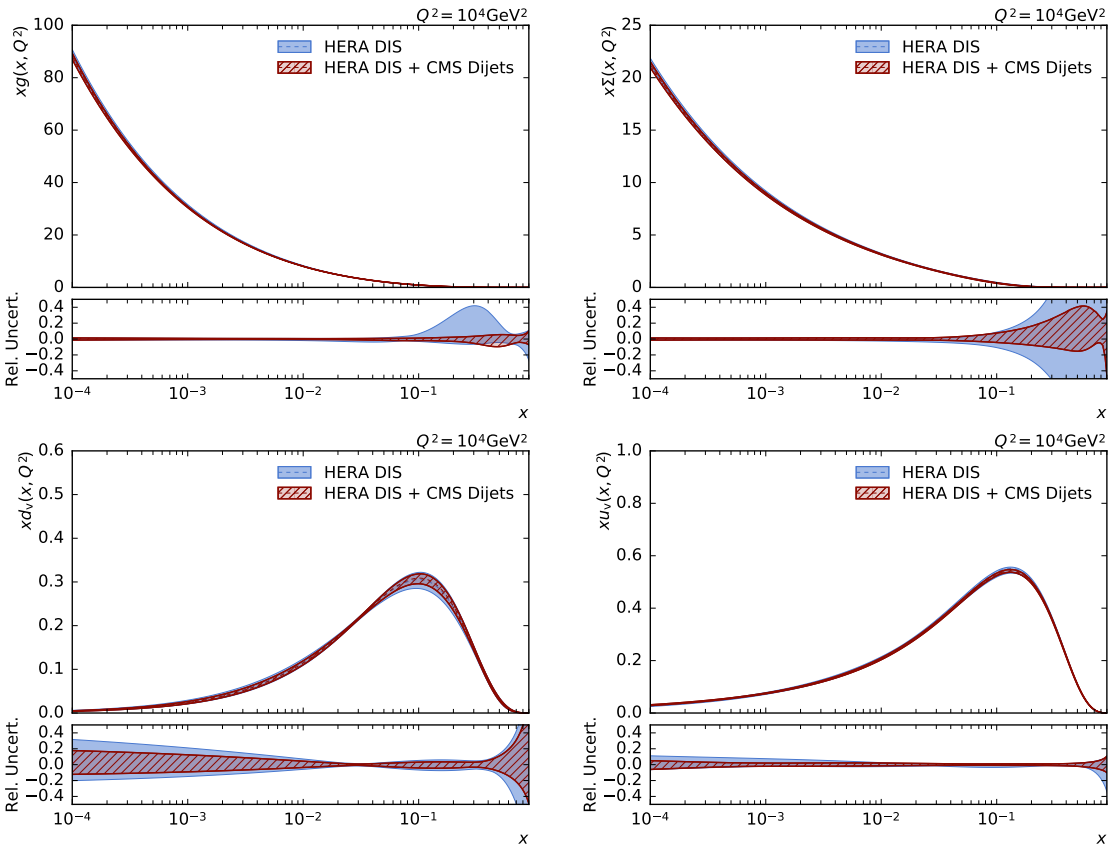

Figure A.14: The gluon (top left), sea quark (top right), d valence quark (bottom left) and u valence quark (bottom right) PDFs as a function of x as derived from HERA inclusive DIS data alone (hatched band) and in combination with CMS dijet data (solid band). The PDFs are evolved to the scale $Q^2 = 10\,000\,\mathrm{GeV}^2$. The total uncertainty of the PDFs is shown.

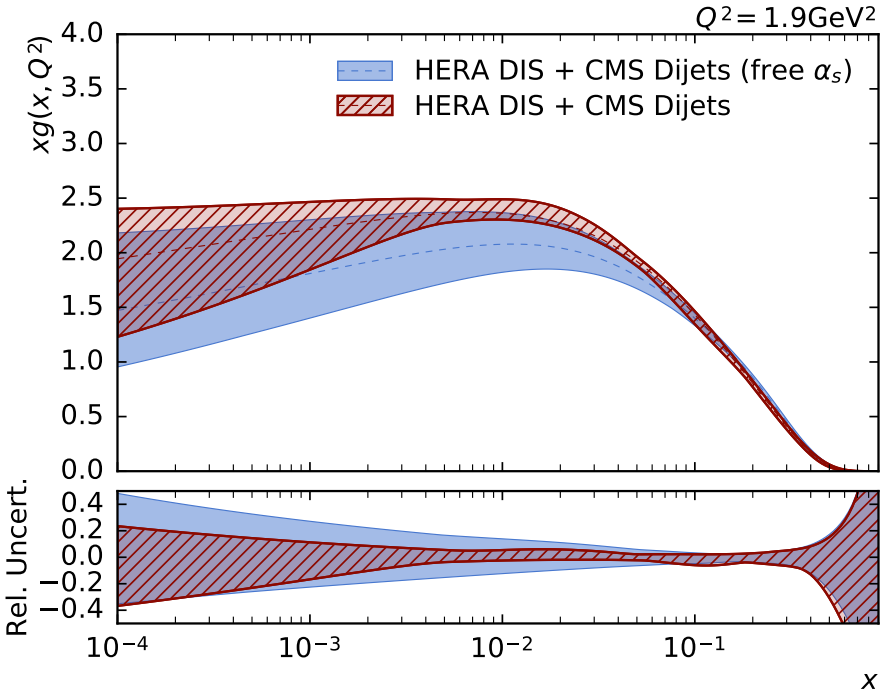

Figure A.15: The gluon PDF as a function of x as derived from HERA inclusive DIS data and CMS dijet data. It is shown with $\alpha_s(M_Z)$ as free parameter (solid band) which is fitted to $\alpha_s(M_Z) = 0.1194$ and with a fixed value of $\alpha_s(M_Z)$ set to 0.1180 (hatched band). The PDF is shown at the scale $Q^2 = 1.9\,\mathrm{GeV}^2$ with total uncertainties.

APPENDIX B

List of Figures

APPENDIX C

List of Tables

APPENDIX D

Bibliography

[1] Michael E. Peskin and Daniel V. Schroeder. „An Introduction to quantum field theory“. Addison-Wesley, 1995. ISBN: 9780201503975.

[2] K. A. Olive et al. „Review of Particle Physics“. *Chin. Phys.* C38 (2014 and 2015 update).
DOI: 10.1088/1674-1137/38/9/090001.

[3] 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, 2003. ISBN: 9780521545891.

[4] Andy Buckley et al. „General-purpose event generators for LHC physics“. *Phys. Rept.* 504 (2011), pp. 145–233.
DOI: 10.1016/j.physrep.2011.03.005. arXiv: 1101.2599 [hep-ph].

[5] Nicola Cabibbo. „Unitary Symmetry and Leptonic Decays“. *Phys. Rev. Lett.* 10 (1963), pp. 531–533.
DOI: 10.1103/PhysRevLett.10.531.

[6] Makoto Kobayashi and Toshihide Maskawa. „CP Violation in the Renormalizable Theory of Weak Interaction“. *Prog. Theor. Phys.* 49 (1973), pp. 652–657.
DOI: 10.1143/PTP.49.652.

[7] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. „Remarks on the unified model of elementary particles“. *Prog. Theor. Phys.* 28 (1962), pp. 870–880.
DOI: 10.1143/PTP.28.870.

[8] B. Pontecorvo. „Inverse beta processes and nonconservation of lepton charge“. *Sov. Phys. JETP* 7 (1958), pp. 172–173.

[9] G. Zweig. „An SU(3) model for strong interaction symmetry and its breaking. Version 2“. *Developments in the Quark Theory of Hadrons, Volume 1.* (1964). Ed. by D.B. Lichtenberg and Simon Peter Rosen, pp. 22–101.

[10] H. Fritzsch, Murray Gell-Mann, and H. Leutwyler. „Advantages of the Color Octet Gluon Picture“. *Phys. Lett.* B47 (1973), pp. 365–368.
DOI: 10.1016/0370-2693(73)90625-4.

[11] CMS Collaboration. „Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC“. *Phys. Lett.* B716 (2012), pp. 30–61.
DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[12] ATLAS Collaboration. „Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC“. *Phys. Lett.* B716 (2012), pp. 1–29.
DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[13] L. D. Faddeev and V. N. Popov. „Feynman Diagrams for the Yang-Mills Field“. *Phys. Lett.* B25 (1967), pp. 29–30.
DOI: 10.1016/0370-2693(67)90067-6.

[14] Steven Weinberg. „New approach to the renormalization group“. *Phys. Rev.* D8 (1973), pp. 3497–3509.
DOI: 10.1103/PhysRevD.8.3497.

[15] Gerard 't Hooft. „Dimensional regularization and the renormalization group“. *Nucl. Phys.* B61 (1973), pp. 455–468.
DOI: 10.1016/0550-3213(73)90376-3.

[16] CMS Collaboration. „Constraints on parton distribution functions and extraction of the strong coupling constant from the inclusive jet cross section in pp collisions at $\sqrt{s} = 7$ TeV“. *Eur. Phys. J.* C75.6 (2015), p. 288.
DOI: 10.1140/epjc/s10052-015-3499-1. arXiv: 1410.6765 [hep-ex].

[17] CMS Collaboration. „Measurement of the inclusive 3-jet production differential cross section in proton–proton collisions at 7 TeV and determination of the strong coupling constant in the TeV range“. *Eur. Phys. J.* C75.5 (2015), p. 186.
DOI: 10.1140/epjc/s10052-015-3376-y. arXiv: 1412.1633 [hep-ex].

[18] R.S. Thorne. „A Variable-flavor number scheme for NNLO“. *Phys.Rev.* D73 (2006).
DOI: 10.1103/PhysRevD.73.054019. arXiv: hep-ph/0601245 [hep-ph].

[19] Sayipjamal Dulat et al. „New parton distribution functions from a global analysis of quantum chromodynamics“. *Phys. Rev.* D93.3 (2016).
DOI: 10.1103/PhysRevD.93.033006. arXiv: 1506.07443 [hep-ph].

[20] L. A. Harland-Lang et al. „Parton distributions in the LHC era: MMHT 2014 PDFs“. *Eur. Phys. J.* C75.5 (2015), p. 204.
DOI: 10.1140/epjc/s10052-015-3397-6. arXiv: 1412.3989 [hep-ph].

[21] NNPDF Collaboration. „Parton distributions for the LHC Run II“. *JHEP* 04 (2015), p. 40.
DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849.

[22] S. Alekhin, J. Blumlein, and S. Moch. „The ABM parton distributions tuned to LHC data“. *Phys. Rev.* D89.5 (2014).
DOI: 10.1103/PhysRevD.89.054028. arXiv: 1310.3059 [hep-ph].

[23] ZEUS Collaboration. „Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data“. *Eur. Phys. J.* C75.12 (2015), p. 580.
DOI: 10.1140/epjc/s10052-015-3710-4. arXiv: 1506.06042 [hep-ex].

[24] S. Alekhin et al. „HERAFitter, Open Source QCD Fit Project“. *Eur. Phys. J.* C75.7 (2015), p. 304.
DOI: 10.1140/epjc/s10052-015-3480-z. arXiv: 1410.4412 [hep-ph].

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

[26] Guido Altarelli and G. Parisi. „Asymptotic Freedom in Parton Language“. *Nucl.Phys.* B126 (1977), p. 298.
DOI: 10.1016/0550-3213(77)90384-4.

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

[28] Torbjorn Sjostrand. „Jet Fragmentation of Nearby Partons“. *Nucl. Phys.* B248 (1984), p. 469.
DOI: 10.1016/0550-3213(84)90607-2.

[29] B. R. Webber. „A QCD Model for Jet Fragmentation Including Soft Gluon Interference“. *Nucl. Phys.* B238 (1984), p. 492.
DOI: 10.1016/0550-3213(84)90333-X.

[30] G. Marchesini and B. R. Webber. „Monte Carlo Simulation of General Hard Processes with Coherent QCD Radiation“. *Nucl. Phys.* B310 (1988), p. 461.
DOI: 10.1016/0550-3213(88)90089-2.

[31] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. „FastJet User Manual“. *Eur. Phys. J.* C72 (2012), p. 1896.
DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[32] Gavin P. Salam. „Towards Jetography“. *Eur. Phys. J.* C67 (2010), pp. 637–686.
DOI: 10.1140/epjc/s10052-010-1314-6. arXiv: 0906.1833 [hep-ph].

[33] S. Catani et al. „New clustering algorithm for multi-jet cross-sections in e+ e- annihilation“. *Phys. Lett.* B269 (1991), pp. 432–438.
DOI: 10.1016/0370-2693(91)90196-W.

[34] S. Catani, B. Webber, and Yu. Dokshitzer. „The K_T clustering algorithm for jets in deep inelastic scattering“. *Nucl. Phys. Proc. Suppl.* 29A (1992), pp. 136–143.
DOI: 10.1016/0920-5632(92)90435-U.

[35] Yuri L. Dokshitzer et al. „Better jet clustering algorithms“. *JHEP* 08 (1997), p. 1.
DOI: 10.1088/1126-6708/1997/08/001. arXiv: hep-ph/9707323 [hep-ph].

[36] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. „The anti-k_t jet clustering algorithm“. *JHEP* 04 (2008), p. 63.
DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[37] Gargamelle Neutrino Collaboration. „Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment“. *Phys. Lett.* B46 (1973), pp. 138–140.
DOI: 10.1016/0370-2693(73)90499-1.

[38] UA1 Collaboration. „Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV“. *Phys. Lett.* B122 (1983), pp. 103–116.
DOI: 10.1016/0370-2693(83)91177-2.

[39] ALICE Collaboration. „The ALICE experiment at the CERN LHC“. *JINST* 3 (2008).
DOI: 10.1088/1748-0221/3/08/S08002.

[40] ATLAS Collaboration. „The ATLAS Experiment at the CERN Large Hadron Collider“. *JINST* 3 (2008), S08003.
DOI: 10.1088/1748-0221/3/08/S08003.

[41] G. L. Bayatian and Chatrchyan. „CMS Physics Technical Design Report Volume I: Detector Performance and Software“. Technical Design Report CMS (2006).

[42] CMS Collaboration. „CMS technical design report, volume II: Physics performance“. *J. Phys.* G34 (2007), pp. 995–1579.
DOI: 10.1088/0954-3899/34/6/S01.

[43] CMS Collaboration. „The CMS experiment at the CERN LHC“. *JINST* 3 (2008).
DOI: 10.1088/1748-0221/3/08/S08004.

[44] LHCb Collaboration. „The LHCb Detector at the LHC“. *JINST* 3 (2008).
DOI: 10.1088/1748-0221/3/08/S08005.

[45] CERN. „The LHC accelerator Complex“. 2016. URL: http://public.web.cern.ch/public/en/Research/AccelComplex-en.html (visited on 01/15/2016).

[46] „CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update“. CMS-PAS-LUM-13-001 (2013).

[47] Simon van der Meer. „Calibration of the effective beam height in the ISR“. CERN-ISR-PO-68-31. ISR-PO-68-31. Geneva: CERN, 1968.

[48] Joram Berger. „Search for the Higgs Boson Produced via Vector-Boson Fusion in the Decay Channel H $\rightarrow \tau\tau$“. PhD thesis. KIT, Karlsruhe, 2014-06-11.

[49] David Barney. „CMS Detector Slice“. URL: http://cds.cern.ch/record/2120661 (visited on 01/20/2016).

[50] CMS Collaboration. „CMS Tracker Layers“. 2011. URL: http://cms.web.cern.ch/news/silicon-strips (visited on 12/10/2015).

[51] A. Rose. „The Level-1 Trigger of the CMS experiment at the LHC and the Super-LHC“. PhD thesis. Imperial Coll., London, 2009.

[52] Fred Stober. „Measurement of the three-jet mass cross-section at $\sqrt{s}$=7 TeV“. PhD thesis. KIT, Karlsruhe.

[53] Georg Fleig. „Dynamic Integration of Cloud Resources into Local Computing Clusters and Calibration of the Jet Energy Scale with the CMS Detector“. PhD thesis. KIT, Karlsruhe, 2016.

[54] CMS Collaboration. „CMS computing: Technical Design Report“. Geneva: CERN, 2005.

[55] Fred Stober, Manuel Zeise, and Joram Berger. „Karlsruhe Advanced Package for Physics Analysis“. 2016. URL: https://github.com/KappaAnalysis/Kappa (visited on 01/25/2016).

[56] Joram Berger et al. „ARTUS - A Framework for Event-based Data Analysis in High Energy Physics“. *arXiv* (2015). arXiv: 1511.00852 [hep-ex].

[57] Fred Stober. „grid-control: The Swiss Army knife of job submission tools“. 2016. URL: https://ekptrac.physik.uni-karlsruhe.de/trac/grid-control (visited on 01/25/2016).

[58] R. Brun and F. Rademakers. „ROOT: An object oriented data analysis framework“. *Nucl. Instrum. Meth.* A389 (1997), pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X.

[59] J. D. Hunter. „Matplotlib: A 2D graphics environment“. *Computing In Science & Engineering* 9.3 (2007), pp. 90–95.

[60] Travis E. Oliphant. „Python for Scientific Computing“. *Computing in Science & Engineering.* 9.3 (May 2007), pp. 10–20. ISSN: 1521-9615. DOI: 10.1109/MCSE.2007.58.

[61] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. „PYTHIA 6.4 Physics and Manual“. *JHEP* 05 (2006), p. 26. DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175 [hep-ph].

[62] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. „A Brief Introduction to PYTHIA 8.1“. *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[63] G. Corcella et al. „HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)“. *JHEP* 01 (2001), p. 10. DOI: 10.1088/1126-6708/2001/01/010. arXiv: hep-ph/0011363 [hep-ph].

[64] M. Bahr et al. „Herwig++ Physics and Manual“. *Eur. Phys. J.* C58 (2008), pp. 639–707. DOI: 10.1140/epjc/s10052-008-0798-9. arXiv: 0803.0883 [hep-ph].

[65] Johannes Bellm et al. „Herwig 7.0 / Herwig++ 3.0 Release Note“ (2015). arXiv: 1512.01178 [hep-ph].

[66] Stefano Frixione, Paolo Nason, and Carlo Oleari. „Matching NLO QCD computations with Parton Shower simulations: the POWHEG method“. *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[67] Paolo Nason. „A New method for combining NLO QCD with shower Monte Carlo algorithms“. *JHEP* 11 (2004), p. 040.
DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146 [hep-ph].

[68] Carlo Oleari. „The POWHEG-BOX“. *Nucl. Phys. Proc. Suppl.* 205-206 (2010), pp. 36–41.
DOI: 10.1016/j.nuclphysbps.2010.08.016. arXiv: 1007.3893 [hep-ph].

[69] Johan Alwall et al. „MadGraph 5: Going Beyond“. *JHEP* 06 (2011), p. 128.
DOI: 10.1007/JHEP06(2011)128. arXiv: 1106.0522 [hep-ph].

[70] Zoltan Nagy. „Next-to-leading order calculation of three jet observables in hadron hadron collision“. *Phys. Rev.* D68 (2003), p. 094002.
DOI: 10.1103/PhysRevD.68.094002. arXiv: hep-ph/0307268 [hep-ph].

[71] T. Kluge, K. Rabbertz, and M. Wobisch. „FastNLO: Fast pQCD calculations for PDF fits“. *Proceedings DIS 2006, 14th International Workshop*. 2006, pp. 483–486. arXiv: hep-ph/0609285 [hep-ph].

[72] Daniel Britzger et al. „New features in version 2 of the fastNLO project“ (2012), p. 217.
DOI: 10.3204/DESY-PROC-2012-02/165. arXiv: 1208.3641 [hep-ph].

[73] M. R. Whalley, D. Bourilkov, and R. C. Group. „The Les Houches accord PDFs (LHAPDF) and LHAGLUE“. *Proceedings: Workshop on the implications of HERA for LHC physics.* 2005. arXiv: hep-ph/0508110 [hep-ph].

[74] Andy Buckley et al. „LHAPDF 6: parton density access in the LHC precision era“. *Eur. Phys. J.* C75 (2015), p. 132.
DOI: 10.1140/epjc/s10052-015-3318-8. arXiv: 1412.7420 [hep-ph].

[75] CMS Collaboration. „Particle–Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET“. CMS Physics Analysis Summary CMS-PAS-PFT-09-001. 2009.

[76] CMS Collaboration. „Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector“. CMS Physics Analysis Summary CMS-PAS-PFT-10-001. 2010.

[77] W. Adam et al. „Track reconstruction in the CMS tracker“ (2005).

[78] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. „The Catchment Area of Jets“. *JHEP* 04 (2008), p. 005.
DOI: 10.1088/1126-6708/2008/04/005. arXiv: 0802.1188 [hep-ph].

[79] CMS Collaboration. „Jet performance in CMS“. *PoS* EPS-HEP2013 (2013), p. 433.

[80] CMS Collaboration Collaboration. „Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS“. *JINST* 6 (2011), P11002.
DOI: 10.1088/1748-0221/6/11/P11002. arXiv: 1107.4277 [physics.ins-det].

[81] CMS Collaboration. „Jet Energy Calibration in the 8 TeV pp data“. *To be published in JINST* (2016).

[82] W. T. Giele, E. W. Nigel Glover, and David A. Kosower. „The inclusive two jet triply differential cross-section“. *Phys. Rev.* D52 (1995), pp. 1486–1499. DOI: 10.1103/PhysRevD.52.1486. arXiv: hep-ph/9412338 [hep-ph].

[83] ATLAS Collaboration. „Measurement of inclusive jet and dijet production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector“. *Phys. Rev.* D86 (2012), p. 014022. DOI: 10.1103/PhysRevD.86.014022. arXiv: 1112.6297 [hep-ex].

[84] Stephen D. Ellis, Zoltan Kunszt, and Davison E. Soper. „Two jet production in hadron collisions at order α_S^3 in QCD“. *Phys. Rev. Lett.* 69 (1992), pp. 1496–1499. DOI: 10.1103/PhysRevLett.69.1496.

[85] Michael H. Seymour and Andrzej Siodmok. „Constraining MPI models using σ_{eff} and recent Tevatron and LHC Underlying Event data“. *JHEP* 10 (2013), p. 113. DOI: 10.1007/JHEP10(2013)113. arXiv: 1307.5015 [hep-ph].

[86] CMS Collaboration. „Event generator tunes obtained from underlying event and multiparton scattering measurements“ (2015). arXiv: 1512.00815 [hep-ex].

[87] M. Cacciari et al. „The t anti-t cross-section at 1.8-TeV and 1.96-TeV: A Study of the systematics due to parton densities and scale dependence“. *JHEP* 04 (2004), p. 068. DOI: 10.1088/1126-6708/2004/04/068. arXiv: hep-ph/0303085 [hep-ph].

[88] Johan Alwall et al. „A Standard format for Les Houches event files“. *Comput. Phys. Commun.* 176 (2007), pp. 300–304. DOI: 10.1016/j.cpc.2006.11.010. arXiv: hep-ph/0609017 [hep-ph].

[89] CMS Collaboration Collaboration. „Recommended MET Filters“. 2012. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMetAnalysis (visited on 08/20/2015).

[90] CMS Collaboration Collaboration. „Jet Identification at 8 TeV“. 2012. URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID (visited on 08/22/2015).

[91] CMS Collaboration Collaboration. „Jet Energy Resolution at 8 TeV“. 2012. URL: https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution (visited on 08/20/2015).

[92] G. D'Agostini. „A Multidimensional unfolding method based on Bayes' theorem“. *Nucl. Instrum. Meth.* A362 (1995), pp. 487–498. DOI: 10.1016/0168-9002(95)00274-X.

[93] Tim Adye. „Unfolding algorithms and tests using RooUnfold“. *Proceedings of the PHYSTAT 2011 Workshop, CERN, Geneva, Switzerland, January 2011, CERN-2011-006, pp 313-318*. 2011, pp. 313–318. arXiv: 1105.1160 [physics.data-an].

[94] Stefan Dittmaier, Alexander Huss, and Christian Speckner. „Weak radiative corrections to dijet production at hadron colliders“. *JHEP* 11 (2012), p. 095. DOI: 10.1007/JHEP11(2012)095. arXiv: 1210.0438 [hep-ph].

[95] M. Botje. „QCDNUM: Fast QCD Evolution and Convolution“. *Comput. Phys. Commun.* 182 (2011), pp. 490–532.
DOI: 10.1016/j.cpc.2010.10.020. arXiv: 1005.1481 [hep-ph].

[96] R.S. Thorne and R.G. Roberts. „An Ordered analysis of heavy flavor production in deep inelastic scattering“. *Phys.Rev.* D57 (1998), pp. 6871–6898.
DOI: 10.1103/PhysRevD.57.6871. arXiv: hep-ph/9709442 [hep-ph].

[97] J. Pumplin et al. „Uncertainties of predictions from parton distribution functions. 2. The Hessian method“. *Phys. Rev.* D65 (2001), p. 014013.
DOI: 10.1103/PhysRevD.65.014013. arXiv: hep-ph/0101032 [hep-ph].

[98] Louis Lyons, Alexander J. Martin, and David H. Saxon. „On the Determination of the B Lifetime by Combining the Results of Different Experiments“. *Phys. Rev.* D41 (1990), p. 982.
DOI: 10.1103/PhysRevD.41.982.

Danksagung

An erster Stelle möchte ich mich bei Prof. Günter Quast für die ausgezeichnete Betreuung bedanken. Die vielen Diskussionen, insbesondere über statistische Methoden, haben entscheidend zum Gelingen dieser Arbeit beigetragen. Ebenso gilt mein Dank Prof. Müller für die Übernahme des Korreferats. Auch seine vielen interessanten Vorlesungen während meines Studiums werde ich immer in guter Erinnerung behalten.

Besonderer Dank gebührt meinem Betreuer Klaus Rabbertz. Sein umfassendes Wissen der Quantenchromodynamik hat erst den Anstoß für diese Arbeit gegeben. Immer wenn ich mich in einem Problem verrannt habe, hatte er den richtigen Tipp parat. Die unzähligen Diskussionen werde ich immer in guter Erinnerung behalten.

Außerdem möchte ich mich bei Mikko Voutilainen für die vielen hilfreichen Kommentare zu dieser Analyse bedanken. Ringaile Placakyte hat mich mit dem HERAFitter Projekt vertraut gemacht und mir bei der Bewältigung der zahlreichen Probleme enorm geholfen.

Ausserdem möchte ich mich bei Fred Stober, Joram Berger und Dominik Haitz für die vielen guten Ideen und hilfreiche Diskussionen bedanken. Die kameradschaftliche Arbeitsatmosphäre am EKP und die vielen lustigen Stunden in der Arbeitsgruppe mit Raphael Friese, Rene Caspart, Thomas Müller und Fabio Colombo haben mir immer sehr gut gefallen.

Zuallerletzt geht mein grösster Dank an meine Eltern für das Vertrauen und die langjährige Unterstützung.

Erklärung der selbständigen Anfertigung der Dissertationsschrift

Hiermit erkläre ich, dass ich die Dissertationsschrift mit dem Titel

»*Measurement of Triple-Differential Dijet Cross Sections with the CMS Detector at 8 TeV and PDF Constraints*«

selbständig und unter ausschließlicher Verwendung der angegebenen Hilfsmittel angefertigt habe.

Georg Sieber
Karlsruhe, den 04. Mai 2016