IEKP-KA/2016-04

Karlsruher Institut für Technologie

Precision Studies of Proton Structure and Jet Energy Scale with the CMS Detector at the LHC

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
bei der Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)
vorgelegte

DISSERTATION

von

Dipl.-Phys. Dominik Haitz
aus Karlsruhe

Tag der mündlichen Prüfung: 20. Mai 2016

Referent: Prof. Dr. Günter Quast
Korreferent: Prof. Dr. Wim de Boer

Contents

1 Introduction

If I have seen further it is by standing on the shoulders of giants.

– Isaac Newton

A single proton in the darkness. Slowly, it begins to move, to accelerate. Invisible forces guide it on a circular path. It joins others of its kind, a pack of particles, round and round it goes, evading a deadly crash again and again and again until – everything is illuminated in the light of the collision.

Science has come a long way. Ages after the bathtub adventures of Archimedes and Galileo's star-gazing, it has evolved into a worldwide endeavour to uncover the secrets of the cosmos. These efforts have led to the construction of increasingly complex experiments: Giant devices probe the laws of nature by smashing particles together and observing the outcome.

Currently, the most powerful and advanced particle accelerator is the Large Hadron Collider (LHC). It is capable of colliding protons at energies and intensities which surpass all previous experiments. This enables the study of physical processes which occur only at high momentum transfers.

To observe these processes, large particle detectors are installed at the interaction points of the LHC. One of these detectors is the Compact Muon Solenoid (CMS). It is designed to measure as many of the particles produced in a collision event as possible, thus providing scientists with extensive information for physics analyses.

These analyses are often difficult: The rare occurrence of certain processes, multiple concurrent proton-proton interactions or the difficult energy measurement and event reconstruction constitute severe experimental challenges. These effects can lead to significant statistical and systematic uncertainties.

The measurement of Z bosons is less affected by some of these problems: The properties of the Z boson are well known from previous experiments; in its dimuon and dielectron decay channels it offers a clear signature with only small contributions from background processes. As a consequence, Z bosons can be precisely measured. At the LHC, Z bosons are produced at an unprecedented rate.

Z+jet events can be utilized for studies of the proton structure. In the parton model, this structure is expressed by the parton distribution functions (PDFs) which predict the probability to find a certain proton constituent (a parton) with a proton momentum fraction x at an energy scale Q. The PDFs are not predicted by perturbative QCD but have to be experimentally determined.

In this thesis, a method to constrain the parameters of the PDFs by measuring the distributions of kinematic quantities of Z bosons is explored. This method exploits the correlation between the PDFs and the expected number of events with Z bosons in particular phase space regions. By fitting the PDFs to these data, the PDF parameters can be determined. If the measurement is precise enough, the uncertainties in the PDFs can be reduced.

Z+jet events can also be used for jet energy calibration: All physics analyses at the LHC rely on the precise reconstruction of the objects produced in a collision. Among the most important of these objects are jets, collimated streams of particles produced by the hadronization of partons. As there are numerous effects that bias the jet measurement, the precise determination of jet energies is among the most challenging experimental tasks.

Sophisticated techniques have been developed to deal with the various systematic biases. One of the most important steps is the data-driven calibration with balancing methods: Exploiting momentum conservation, the jet transverse momentum is compared with the transverse momentum of a well-measured reference object and consequently corrected. In this thesis, the jet energy scale is calibrated by studying $Z(\rightarrow \mu^+\mu^-)$+jet events.

Outline

In Chapter 2 of this thesis, the theoretical foundations for the performed studies are described. A brief introduction to quantum chromodynamics is followed by an explanation of the PDFs and the production and decay of Z bosons.

Chapter 3 discusses the experimental setup. The LHC and the CMS detector with its different components are described. It is explained how the signals from these components are used to reconstruct physical objects, e.g. particles or jets.

In Chapter 4, the software tools utilized for data analysis, Monte Carlo simulation and PDF determination are outlined.

The calibration studies with $Z(\rightarrow \mu^+\mu^-)$+jet events are described in Chapter 5. First, an overview of the jet correction procedure in the CMS Collaboration is given. This is followed by a description of the $Z(\rightarrow \mu^+\mu^-)$+jet event selection and analysis procedure. Several studies and cross-checks of jet and particle kinematic quantities, jet response, jet energy resolution and jet flavour composition are performed. Finally, the absolute corrections on the jet energy scale are determined.

Chapter 6 describes the analysis of $Z(\rightarrow e^+e^-)$+jet events and the determination of the PDFs. The event selection, the estimation of contributions from background processes, and the unfolding of the measured distributions are outlined. The experimental and theoretical uncertainties are estimated. The results are compared to theory predictions for different PDFs. A fit of the PDFs to the measured data is performed and constraints on the PDFs are extracted.

The thesis concludes with Chapter 7, summarizing how Z+jet studies are pushing the precision frontier.

2 Theoretical Foundations

The work carried out in this thesis relies upon the latest developments in theoretical particle physics: The theory of quantum chromodynamics is necessary to understand the interactions between partons and the formation of jets. Parton distribution functions describe the structure of hadrons and are especially important for cross section calculations in proton-proton collisions. Z bosons can be produced in hadron collisions and provide an excellent opportunity for calibration because of their well-known properties.

This chapter will give an overview of the relevant aspects of these topics. General information on theoretical particle physics or the Standard Model can be found in [1–3]. The unit convention $\hbar = c = 1$ is used throughout this thesis.

2.1 Phenomenology of Quantum Chromodynamics

Quantum chromodynamics (QCD) is a quantum field theory describing the strong interaction. It was initially introduced to explain the existence of the Δ^{++} particle: The quark model successfully describes hadrons as consisting of different combinations of elementary particles called quarks. In this model, the Δ^{++} particle is made up of three up quarks with the same spin state. With its symmetric wave function, this particle seemingly violates the Pauli exclusion principle.

As a solution, a new quantum number to re-establish the necessary anti-symmetry was postulated: the *colour charge* with the basic values red, green, blue and the corresponding anti-colours. Colour charge is described as a quantum number under SU(3) symmetry.

This gives rise to a new group of particles: gluons, massless vector bosons which mediate the strong force between quarks. Group theory predicts a singlet and an octet gluon state. As SU(3) is non-abelian, the eight gluons of the octet carry colour charge themselves and therefore also couple to each other. This gluon self-coupling limits the range of the strong force.

The gluon singlet state would be colour-neutral. This would result in an infinite range of the strong force, a behaviour which is not observed in nature. Thus, the singlet-state gluon is postulated not to exist.

The strong force has a peculiar behaviour: The value of its coupling "constant" $\alpha_s(Q^2)$ decreases with the energy scale Q^2 of the interaction. This behaviour is known as the *running of the strong coupling*. This is in contrast to the electromagnetic interaction where the coupling constant α shows only a slight energy-dependence. The energy dependence of $\alpha_s(Q^2)$ can be calculated if the constant has been experimentally determined at one value of Q^2. The running of the strong coupling has been confirmed in various measurements at different energies, see Figure 2.1.

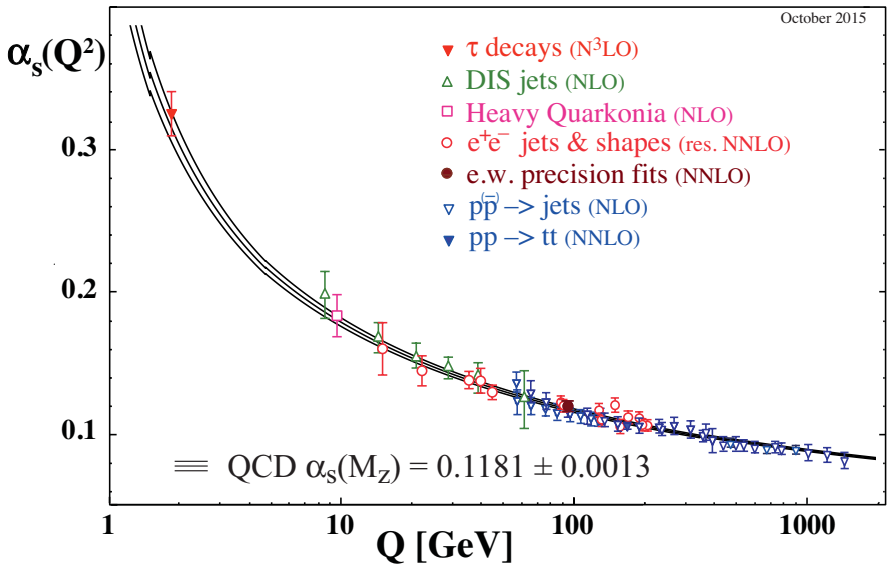


Figure 2.1: Running of the strong coupling. Recent measurements performed by the CMS Collaboration [4, 5] confirm the predicted behaviour at highest energies [6].

With the running of the strong coupling, QCD is able to explain two phenomena: confinement and asymptotic freedom.

Confinement Confinement denotes the observation that colour-charged particles like quarks and gluons are not observed as isolated particles in nature. Instead, they are always part of bound objects such as hadrons which are colour-neutral, e.g. consist of a combination of a red, blue and green quark.

If two colour-charged objects are pulled apart, the strong force acting between them increases. This interaction is mediated via gluons, which carry colour charge and couple to themselves: A gluon "string" is created between the two objects. Because of the coupling strength increasing with distance, enough energy is accumulated to create quark-antiquark pairs out of the vacuum.

In high-energetic hadron collisions where partons are freed from the colliding hadrons, this process repeatedly occurs until the collision energy is used up. Subsequently, the created quarks and antiquarks form hadrons. This leads to a final state of collimated, colourless hadronic particles – a particle jet.

Asymptotic freedom Contrary to confinement, asymptotic freedom describes the behaviour of the strong coupling at high energies or short distances. Here, α_s becomes small and quarks can be treated as quasi-free particles. The low value of α_s enables the application of perturbation theory. In conjunction with a model for describing the structure of hadrons, cross sections in hadron collisions can be calculated.

2.1.1 The QCD Lagrangian

The Lagrangian of QCD is [7]

$$\mathcal{L} = \sum_f \bar{\psi}_{i,f}(i\gamma_\mu D^\mu_{ij} - m_f\delta_{ij})\psi_{j,f} - \frac{1}{4}F^a_{\mu\nu}F^{a\,\mu\nu} \tag{2.1}$$

$$\text{with} \quad F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g_s f_{abc} A^b_\mu A^c_\nu \tag{2.2}$$

$$\text{and} \quad D^\mu_{i,j} = \partial_\mu \delta_{ij} + i g_s t^a_{ij} A^\mu_a \,. \tag{2.3}$$

Repeated indices are summed over. The indices a, b and c refer to colour indices for the adjoint representation of SU(3), i.e. $a, b, c = 1, ..., 8$ for the eight gluons. For the fundamental representation of SU(3), the indices $i, j = 1, ..., 3$ represent the three colours. f corresponds to quark flavour and A to the vector potential of the gluon field. The t^a_{ij} are the generators of SU(3). γ_μ denote the Dirac matrices.

The first term of the Lagrangian with the covariant derivative D describes a free quark and its interaction with the gluon field. As SU(3) is non-abelian, the structure constant f in the field strength tensor F is non-vanishing which leads to gluon self-interactions (3-gluon and 4-gluon vertices). The three fundamental QCD vertices are shown in Figure 2.2. All interactions depend on the coupling $g_s = \sqrt{4\pi\alpha_s}$. The complete set of Feynman rules derived from the Lagrangian can be found in e.g. [8].

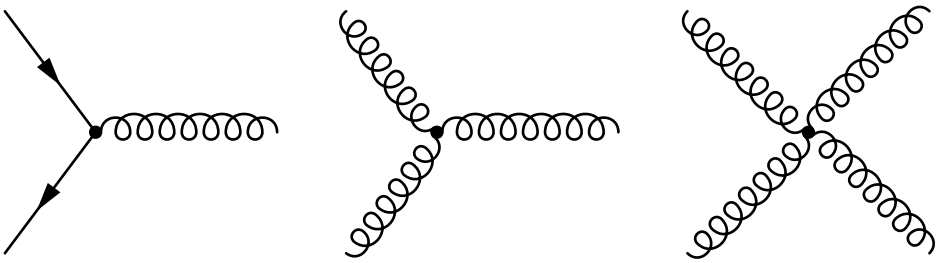


Figure 2.2: The three fundamental QCD interaction vertices.

2.1.2 The Parton Model

In the parton model, hadrons are described as consisting of quarks and gluons, collectively referred to as partons, which carry fractions of the hadron momentum. In QCD, this is described by the parton distribution functions (PDFs). They represent the probability to find a parton with a momentum fraction x at an energy scale Q.

The x-dependence of the PDFs cannot be perturbatively calculated. Instead, the PDFs are parametrized and the correlation between PDFs and observables such as cross sections in hadron collisions is exploited to constrain the PDF parameters.

Such studies have been extensively performed in deep-inelastic scattering experiments at the HERA $e^{\pm}p$ collider [9]. In electron-proton scattering, the kinematic quantities of the scattered electron are measured to determine the PDFs. Exemplary Feynman diagrams of charged and neutral current interactions are shown in Figure 2.3.

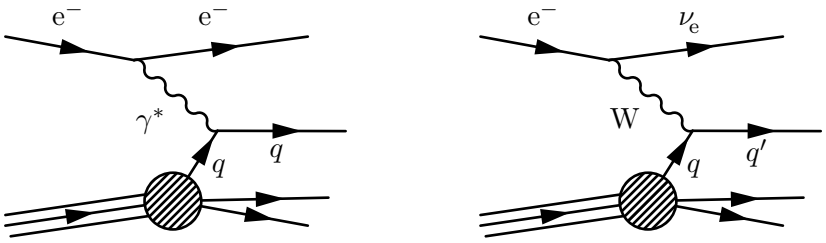


Figure 2.3: Feynman diagrams for neutral current (left) and charged current (right) interactions in deep-inelastic scattering.

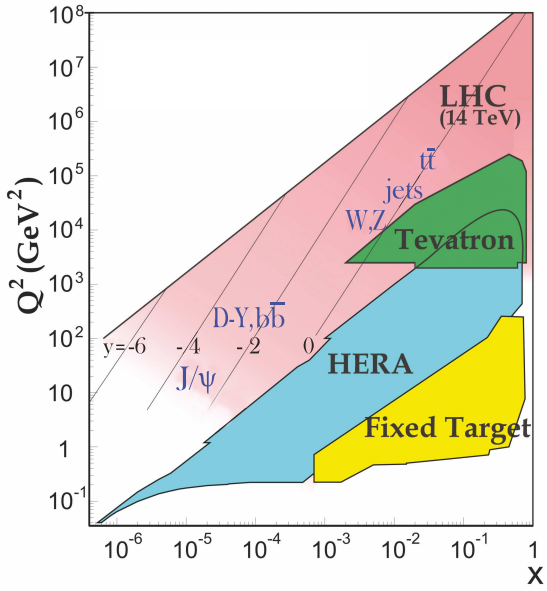


Figure 2.4: Accessible phase space in x and Q^2 at the LHC compared to HERA, Tevatron and fixed target experiments [6].

Further knowledge can be derived by increasing the experimental precision or the accessible phase space. At the LHC, extended coverage in x and Q is achieved compared to HERA, see Figure 2.4. PDF constraints were extracted from e.g. inclusive jet measurements [4, 10].

The most precise PDFs are produced by analysis groups which take multiple measurements into account, e.g. the NNPDF [11], HERAPDF [9], ABM [12], CTEQ [13] and MMHT [14] collaborations. They differ in the fitting method, the included measurements, the PDF parametrization or the treatment of uncertainties. An example PDF from the NNPDF collaboration can be seen in Figure 2.5.

While the x-dependence of the PDFs has to be experimentally determined, the dependence on Q can be calculated with specific equations, described in the following paragraphs.

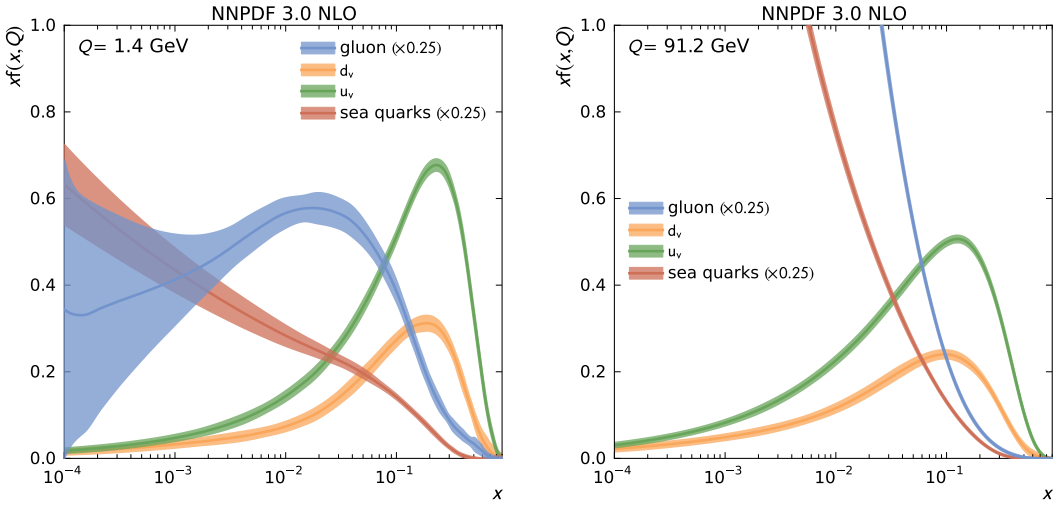


Figure 2.5: PDFs from the NNPDF collaboration at different energy scales of $Q = 1.4\,\text{GeV}$ (left) and $Q = 91.2\,\text{GeV}$ (right). The values for the gluon and sea quarks are scaled by 1/10 to be in a comparable range with the valence quark PDFs. The PDFs are evolved to higher energy scales via the DGLAP equations. At higher Q, the gluon and sea quarks PDFs increase, especially at lower x values.

The DGLAP Equations The energy-dependence of PDFs is due to a higher resolution when probing the proton at higher Q. As partons continuously interact with each other, emission and splitting processes take place. These are described by the splitting functions P_{ab}, where a, b are quarks or gluons. For example, $P_{qg}(x/y)$ describes the probability of a quark with momentum fraction x to originate from a gluon with higher momentum fraction $y > x$.

The splitting functions can be perturbatively calculated as

$$P_{ab}(x, \alpha_s(Q^2)) = P^0_{ab}(x) + \frac{\alpha_s(Q^2)}{2\pi} P^1_{ab}(x) + \mathcal{O}\left(\alpha_s^2(Q^2)\right) . \tag{2.4}$$

They are used by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [15–17] to describe the evolution of PDFs from an energy scale Q_0 to a scale $Q > Q_0$:

$$\frac{\mathrm{d}\, q(x, Q^2)}{\mathrm{d} \ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{\mathrm{d}\, y}{y} \left[P_{qq}\left(\frac{x}{y}\right) \cdot q(y, Q^2) + P_{qg}\left(\frac{x}{y}\right) \cdot g(y, Q^2) \right] , \tag{2.5}$$

$$\frac{\mathrm{d}\, g(x, Q^2)}{\mathrm{d} \ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{\mathrm{d}\, y}{y} \left[\sum_q P_{gq}\left(\frac{x}{y}\right) \cdot q(y, Q^2) + P_{gg}\left(\frac{x}{y}\right) \cdot g(y, Q^2) \right] . \tag{2.6}$$

2.1.3 Perturbative QCD

One of the most important applications of QCD (and an opportunity for stringent tests of the theory) is the calculation of cross sections in hadron collisions. Usually, this is performed via a perturbative approach, leading to the field of perturbative QCD (pQCD).

The small value of α_s at high energies (asymptotic freedom) enables the application of perturbative methods: A theory prediction for an observable X can be expanded in powers of α_s as

$$X = \tilde{X}_1 + \alpha_s \tilde{X}_2 + \alpha_s^2 \tilde{X}_3 + ... \tag{2.7}$$

where a calculation of the first few order is usually sufficient.

This approach is only possible if α_s is small ($\ll 1$), i.e. in the high-energy regime. However, a complete cross section calculation also includes the description of low-energetic interactions. Therefore, it is necessary to factorize the problem into a short-distance, process-specific part (which can be solved with perturbative methods) and a long-distance, universal part (described by phenomenological models).

However, this approach leads to ultraviolet divergences in the cross section calculation. These can be countered via a procedure called *renormalization*: All divergences are absorbed via a redefinition of masses, fields and coupling constants. This introduces the *renormalization scale* μ_r, effectively a momentum cut-off, which can be arbitrarily chosen.

The non-perturbative part of the calculation can be based on the PDFs. As the naive parton model leads to infrared and collinear divergences in the calculation, the *QCD improved parton model* is introduced: In a procedure similar to renormalization, the PDFs absorb all non-perturbative parts of the cross section calculation. This introduces the *factorization scale* μ_f.

Finally, the QCD factorized cross section for a particular process is given as

$$\sigma = \sum_{i,j} \iint \mathrm{d}\, x_1 \, \mathrm{d}\, x_2 \, f_i(x_1, \mu_f) f_j(x_2, \mu_f) \times \sigma_{ij}(x_1, x_2, \alpha_s(\mu_r), \mu_r, \mu_f) \tag{2.8}$$

where the sum includes all partons i, j in the two protons.

Applying perturbation theory yields

$$\sigma_{ij}(x_1, x_2, \alpha_s(\mu_r), \mu_r, \mu_f) = \tilde{\sigma}_1(x_1, x_2, \mu_r, \mu_f) + \alpha_s(\mu_r)\tilde{\sigma}_2(x_1, x_2, \mu_r, \mu_f) + ...$$

Both μ_r and μ_f are arbitrary and are typically set to the energy scale of the hard process. The calculated cross section does not depend on μ_f and μ_r *unless* the perturbative series is truncated, which is usually the case – the practicability of calculating only the first few terms is the reason why perturbation theory was applied in the first place. For perturbative calculations, this introduces a *scale uncertainty* due to the arbitrary choice of μ_r and μ_f. In the example of a process independent of α_s at leading-order (LO), the cross section at next-to-leading-order (NLO) accuracy is calculated as

$$\sigma_{\mathrm{NLO}}(\mu_r, \mu_f) = \sum_{i,j} \iint \mathrm{d}\, x_1 \, \mathrm{d}\, x_2 f_i(x_1, \mu_f) f_j(x_2, \mu_f) \times \left(\tilde{\sigma}_{\mathrm{LO}}(x_1, x_2, \mu_r, \mu_f) + \alpha_s(\mu_r)\tilde{\sigma}_{\mathrm{NLO}}(x_1, x_2, \mu_r, \mu_f)\right).$$

At NLO, the scale dependence is reduced compared to LO since α_s (μ_r) partly compensates the μ_r-dependence of $\sigma(\mu_r, \mu_f)$. The dependence decreases even further at next-to-next-to-leading-order (NNLO). Therefore, at higher orders the cross section calculation does not only become more accurate but also the scale uncertainty is reduced.

Matrix Element Calculation

The probability to go from an initial state i to a final state f, i.e. the transition amplitude, is given via the corresponding element of the scattering matrix $\mathbf{S}$. To calculate the cross section for a specific process $|i\rangle \to |f\rangle$, the squared matrix element $|\mathbf{S}_{fi}|^2$ is integrated over the phase space of the outgoing particles.

In the perturbative expansion, the LO matrix element is represented by the tree-level Feynman diagrams. At NLO, the higher order in α_s leads to additional real and virtual contributions. The real contributions correspond to additional outgoing partons, i.e. external lines in the Feynman diagram. Virtual contributions correspond to loops, i.e. internal lines, in the Feynman diagrams.

For example, the NLO calculation of a $2 \to n$ process includes the LO $2 \to n$ calculation, the $2 \to n+1$ final state (which corresponds to $2 \to n+1$ at LO) and the $2 \to n$ loop corrections.

The cross sections for the most important processes in proton-proton collisions have been determined at NLO or even NNLO accuracy [18]. The PDFs are an important ingredient for this calculation. For many processes (e.g. $gg \to H$), the PDF uncertainty is the dominating contribution to the total theoretical uncertainty. This illustrates the need for further PDF precision measurements. The importance and usage of PDFs at the LHC is described in [19].

Parton Shower and Hadronization

Following the calculation of the cross section, a complete description of the final state includes the evolution of partons emerging from the hard process. This is necessary to compare theory predictions with jet measurements. Parton shower techniques and hadronization models are used for this task. The different stages of a collision event are visualized in Figure 2.6.

Parton shower Colour-charged particles such as partons can radiate gluons. Given the high particle multiplicity in hadron collisions, the description of all emissions via Feynman diagrams is unfeasible.

Instead, the parton shower technique is used. It describes soft and collinear emissions of partons evolving from high to low energy scales. The splitting functions, see Equation 2.4 are used to derive the *Sudakov form factor*:

$$\Delta(Q^2, q^2) = \exp\left[-\int_{q^2}^{Q^2} \mathrm{d}\, k^2 \Gamma(k^2)\right] \tag{2.9}$$

where $\Gamma(k^2)$ is the integrated splitting function. It describes the probability for a parton to go from a scale Q to a lower energy scale q without gluon radiation.

Based on the Sudakov form factor, the emission and splitting probabilities for partons at different energies can be calculated. This is done iteratively until a lower energy threshold, usually around 1 GeV, is reached.

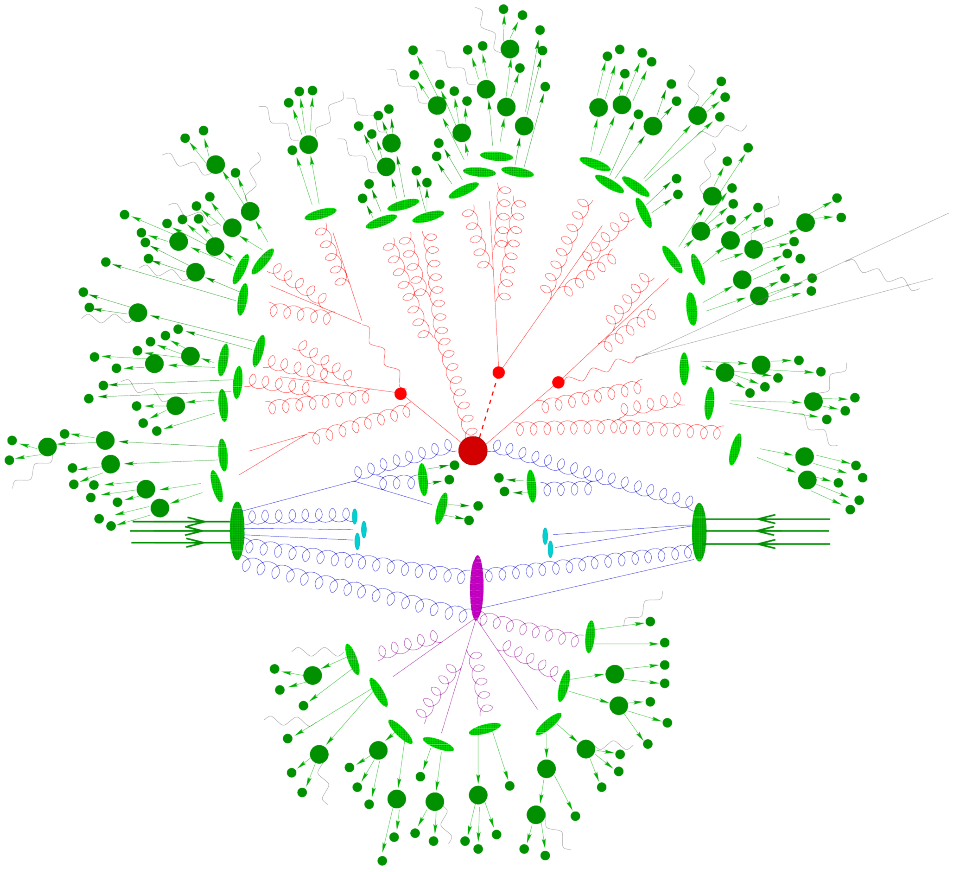


Figure 2.6: Schematic of the different stages of a hard scattering process in a hadron-hadron collision, resulting in a high particle multiplicity in the final state: hard interaction (dark red), parton shower (light red), secondary hard interaction (purple), beam remnant (cyan), hadronization (light green), hadron decays (dark green), soft photon radiation (grey) [20].

While the parton shower accounts for soft or collinear emissions, high-energetic radiation is better described as part of the higher-order Feynman diagrams. To avoid double counting in overlapping phase space regions, the fixed-order calculation has to be matched to the parton shower. This is performed via matching methods such as CKKW [21] or MLM [22].

Hadronization models After the parton showering, stable hadrons are formed in a process called hadronization. This takes place in the low-energy regime where perturbative approaches are not possible. Instead, phenomenological models are employed for the description of hadronization processes.

The Lund string model [23] assumes coloured "strings" between the partons. If enough energy is accumulated, a string "snaps" and a quark-antiquark pair is created out of the vacuum. Subsequently, hadrons are formed from the generated quark pairs.

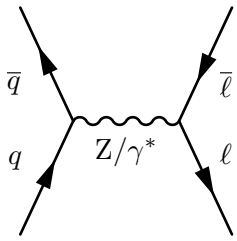


Figure 2.7: Feynman diagram of the Drell-Yan process. Quark-antiquark annihilation gives rise to a Z boson or a virtual photon which decays into two leptons.

2.2 Z Bosons at Hadron Colliders

The Z boson is the neutral exchange particle of the weak interaction. In the 1960s, Glashow, Weinberg and Salam unified the theories of the weak and electromagnetic interactions into the electroweak theory. This theory predicted the existence of the massive neutral Z boson which had not been observed before. In 1983, the Z boson was discovered by the UA1 and UA2 experiments at the Super Proton Synchrotron [24, 25]. In the 1990s, the mass and decay width of the Z boson were measured with high precision at the LEP collider [26]. The most important properties of the Z boson are given in Table 2.1.

At the LHC, Z bosons can be produced via the Drell-Yan process [27], see Figure 2.7. Quark-antiquark annihilation produces a Z boson or a virtual photon which decays into a pair of oppositely charged leptons. For the invariant mass of the dilepton system being close to the Z boson mass, the production of Z bosons is predominant over photon exchange.

Z bosons decay mostly hadronically, see Figure 2.9. The decay into two oppositely-charged leptons of the same flavour occurs with a branching ratio of only around 3.3% per flavour ($\mu^+\mu^-$, $\mathrm{e}^+\mathrm{e}^-$ or $\tau^+\tau^-$). However, because of its clear signature, the Z boson decay into a dielectron or dimuon pair is important for many physics analyses.

Table 2.1: Most important Z boson properties [6].

Property	Value
Charge	0
Spin	1
Mass	(91.1876 ± 0.0021) GeV
Decay width	(2.4952 ± 0.0023) GeV
Average lifetime	3×10^{-25} s

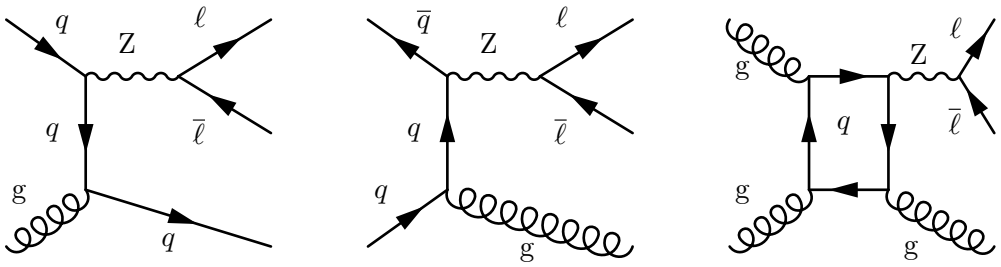


Figure 2.8: Feynman diagrams of Z($\to \bar{\ell}\ell$)+jet production in the gq (left), $q\bar{q}$ (centre) and gg (right) production channels.

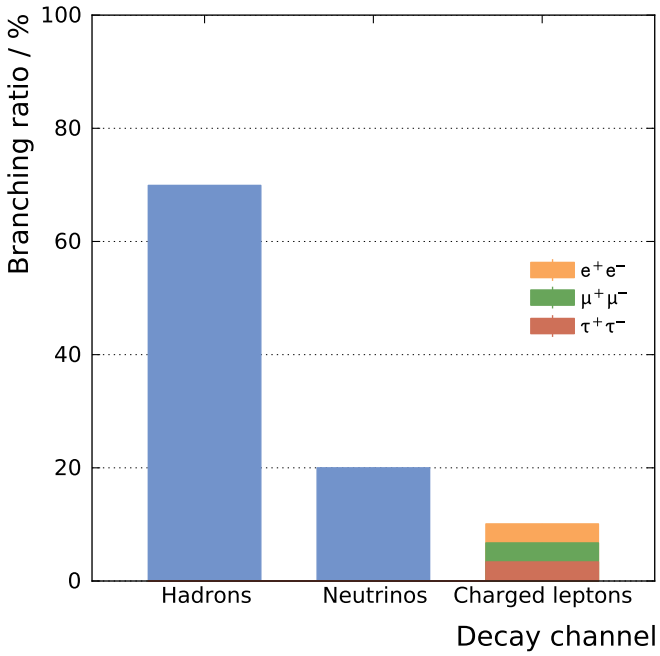


Figure 2.9: Dominant Z boson decay channels and measured branching ratios [6].

3 Experimental Setup

Experimental physics strives towards verifying predictions from theory or observing unknown phenomena. As science progresses, theories have to be tested in new ways or under exceptional circumstances.

In particle physics, this means studying interactions between high-energetic particles. This necessitates the construction of devices to observe and, in a laboratory environment, to induce particle collisions. As particle energies become higher and higher, these experiments become increasingly complex.

At the moment, the most powerful and sophisticated particle accelerator is the LHC (Large Hadron Collider) [28] at CERN[1] in Geneva, Switzerland. It is accompanied by a number of experiments, among which is the CMS (Compact Muon Solenoid) detector which recorded the data this thesis is based upon.

With its unprecedented luminosity and centre-of-mass energy, the LHC is pushing the high-energy frontier. In 2012, the analysis of LHC collision data led to the discovery of the Higgs boson [29, 30], a historic milestone in particle physics.

In the following sections, the LHC and its most important characteristics, referring to the 2012 operational period, are briefly described. The CMS detector is explained in more detail: An overview of its hardware components is followed by a description of how the measured signals are used to reconstruct physical objects, e.g. individual particles or particle jets.

3.1 The Large Hadron Collider

Particle collision experiments have been conducted since the 1950s, with varying technical configuration (circular or linear shape) or different particles (electron-positron, proton-antiproton or proton-proton colliders). The circular LHC with its two opposite-directed beams was primarily built for proton-proton collisions.

This design choice takes several considerations into account:

- The main advantage of using protons compared to the lighter electrons is the much lower synchrotron radiation: As the radiative energy loss is proportional to $1/m^4$, an electron collider of the same size could by far not achieve the LHC's collision energies.

- While proton-antiproton colliders are in some aspects easier to construct (the same electric and magnetic fields can accelerate and steer protons and the opposite-

[1] European Organization for Nuclear Research / Organisation Européenne pour la Recherche Nucléaire

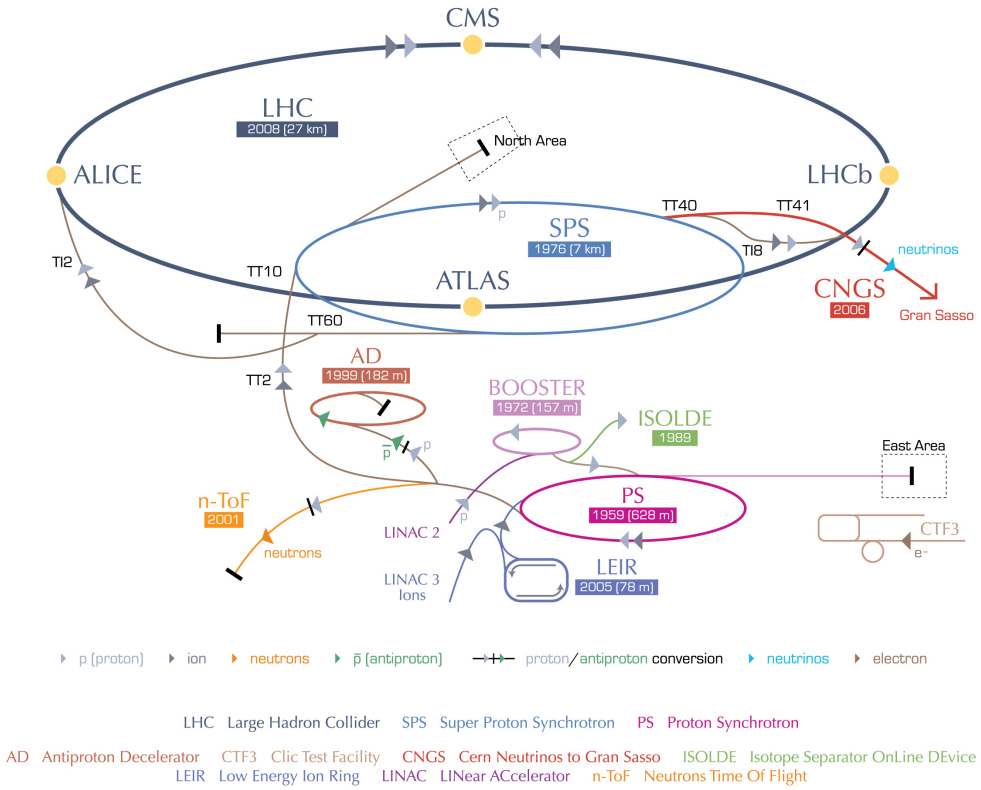


Figure 3.1: An overview of the LHC accelerator chain. Protons are accelerated by various small facilities before they enter the LHC ring [32].

directed antiprotons at the same time), antiprotons can at present not be produced at the rate needed for the collision frequency of the LHC.

- The circular shape allows the beams to be used over and over again, in contrast to a linear collider, where the beams are discarded after each collision.

The most important technical parameters of a particle collider are the centre-of-mass energy of the collision and the luminosity of the beams. A high centre-of-mass energy is needed to produce heavy particles, a high luminosity enables the observation of rare processes. The LHC surpasses previous collider experiments in both aspects: In 2012, a centre-of-mass energy of 8 TeV and a maximum luminosity of $7.67 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ were reached.

The LHC was installed in the 27 km long tunnel previously occupied by its predecessor LEP[2] [31], 50 m to 175 m below the surface. It passes under the main CERN site in Meyrin, 8 km west of Geneva, and crosses the border into France.

[2]Large Electron Positron Collider

Protons are accelerated by several smaller facilities before being injected into the main ring. An overview of the LHC with its pre-accelerator chain is given in Figure 3.1. Having reached the LHC pipe, the protons are further accelerated in 16 cavities by strong electric fields. 1232 dipole magnets, each generating a field of 8.33 T, force the protons on a circular trajectory. 392 quadrupole magnets focus the beams as the proton bunches spread out because of the magnetic repulsion between the protons. The magnets are superconducting and are therefore operated at a temperature of 2 K.

As protons in the LHC are accelerated via electromagnetic waves, the beams cannot consist of a continuous stream. Instead, the protons have to be packed into bunches. A proton beam is divided into 1380 discrete bunches, each containing around $1.15 \cdot 10^{11}$ protons. The time difference between two bunches is 50 ns.

The two beams cross at four interaction points, synchronized such that the bunches pass the interaction points at the same time. To observe the resulting particle collisions, several large particle detectors are installed at the interaction points, see Figure 3.1: CMS [33], ATLAS[3] [34] , ALICE[4] [35] and LHCb[5] [36]. The smaller TOTEM[6] [37] and LHCf[7] [38] experiments are installed in the caverns of the CMS and ATLAS detectors, respectively, and complement the physics programme at the LHC by focusing on particles with high rapidities. Even though the different experiments focus on different research topics, in some cases they can perform similar analyses and cross-check each other's results.

The commissioning of the LHC was finished in 2008. However, shortly after the start-up in September that year, a faulty electrical connection lead to a magnet quench and subsequent helium loss [39]. The damage assessment and repairs delayed operation until the end of 2009.

In 2010 and 2011, proton-proton collisions were produced at $\sqrt{s} = 7\,\text{TeV}$. After increasing the centre-of-mass energy to 8 TeV, $23.30\,\text{fb}^{-1}$ of proton-proton collisions were delivered in 2012 [40]. The operational period up to this year is called the *Run 1* of the LHC.

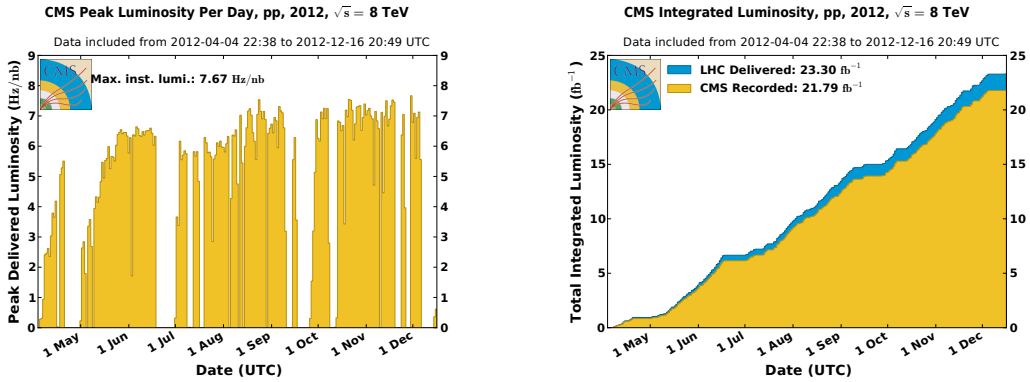
In spring 2013, the LHC and the detectors went into a two-year period for maintenance and upgrade. In 2015, the LHC has resumed operation with proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$. This marked the beginning of *Run 2* of the LHC. For the future, plans exist for further upgrades such as the High Luminosity LHC [41].

Luminosity

In particle collisions, the expected event rate $\partial N/\partial t$ for a particular process is connected to its cross section σ via the luminosity L:

$$\frac{\partial N}{\partial t} = L \cdot \sigma \,. \tag{3.1}$$

[3] A Toroidal LHC ApparatuS
[4] A Large Ion Collider Experiment
[5] Large Hadron Collider beauty experiment
[6] Total Elastic and Diffractive Cross Section Measurement
[7] Large Hadron Collider forward



(a) (Instantaneous) Luminosity

(b) Integrated Luminosity

Figure 3.2: Measured luminosity for LHC proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$. The average luminosity increased over time (left). As not all collisions could be recorded by the CMS detector, the recorded luminosity is slightly lower compared to the delivered one (right).

The total number of expected events for a certain process can be calculated by integrating the event rate over time:

$$N = \int L \cdot \sigma \, \mathrm{d}t = \sigma \cdot \int L \, \mathrm{d}t = \sigma \cdot \mathcal{L} \tag{3.2}$$

where $\mathcal{L}$ denotes the integrated luminosity.

The luminosity is a machine parameter. Its value depends on the particular collider and the operation conditions at certain point in time. At the LHC, given a Gaussian proton distribution in the bunches, the machine luminosity can be expressed as [28]

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi \epsilon_n \beta^* F} \,. \tag{3.3}$$

In this formula, N_b denotes the number of particles in each bunch, n_b is the number of bunches per beam, f_{rev} the beam revolution frequency, ϵ_n the normalized transverse beam emittance and β^* the beta function at the collision point. γ_r denotes the relativistic gamma-factor and F is the geometric luminosity reduction factor.

Figure 3.2(a) shows the development of the luminosity for the collisions recorded with the CMS detector in 2012. The luminosity increased on average, its maximum value was $7.67 \times 10^{33}\,\text{cm}^{-2}\,\text{s}^{-1}$. The integrated luminosity can be seen in Figure 3.2(b), amounting to $23.3\,\text{fb}^{-1}$ for the entire data-taking period.

Pileup Collisions

For physics analyses, a high number of events of a certain process is desired in order to reduce statistical uncertainties. This demand can be met via a high machine luminosity (see Equation 3.2). To achieve a reasonably high luminosity at the LHC, $1.15 \cdot 10^{11}$ protons are packed into one bunch.

Because of this large number of protons, multiple soft proton-proton interactions usually occur during each bunch crossing. These soft collisions are called *pileup*. For the collision data recorded at $\sqrt{s} = 8\,\text{TeV}$, on average 21 interactions per bunch crossing were recorded [40].

Pileup collisions usually result in low energetic particles. These are measured by the detector and overlay the experimental signature of a hard scattering process relevant for physics studies. For analysing the collision data, the separation of particles from pileup and particles from the hard scatter poses a significant experimental challenge. How the CMS Collaboration counters this challenge is explained in Section 3.2.2.

3.2 The Compact Muon Solenoid Detector

The CMS experiment is a large general-purpose detector located at one of the collision points at the LHC. The CMS Collaboration comprises around 4300 active members [42] engaged in a wide range of physics analyses: the measurement of parameters of the Standard Model or the PDFs of the proton, testing the predictions of the Standard Model at previously unreached centre-of-mass energies or the search for physics beyond the Standard Model. One of the collaboration's most prominent successes was the discovery of the Higgs boson in 2012 [29].

Certain technical requirements result from these tasks and from the parameters of the LHC beams: The variety of physics analyses requires the detection of different particle types over a wide range in phase space. This necessitates a design comprising several subdetectors, resulting in a length of 21 m, a diameter of 15 m and a weight of 14 000 t. To detect as many of the particles produced in a collision as possible, i.e. achieve a large geometric coverage, the CMS detector consists of a central (barrel) section complemented by endcap parts on both sides. Additional (*forward*) components are installed on both sides of the endcaps close to the beam pipe.

The CMS detector is comprised of the following subsystems, onion-like arranged in layers from the inside (close to the interaction point) to the outside:

Silicon-based tracking system to determine the trajectories and the momenta of charged particles. A high granularity of the tracker is needed to distinguish particle tracks from primary and secondary vertices, which is important e.g. for the identification of jets originating from b quarks, which occur in Higgs boson or top quark decays.

Electromagnetic calorimeter to measure the energies of photons and electrons. A high spatial and energetic resolution is needed to precisely measure e.g. photons from $\text{H} \to \gamma\gamma$ decays or electrons from $\text{Z} \to \text{e}^+\text{e}^-$ decays.

Hadronic calorimeter to measure the energies of hadrons. A high energy resolution is needed for jet measurements.

Solenoid to create a magnetic field necessary to bend the trajectories of charged particles, necessary for their momentum measurement.

Muon chambers for the identification and measurement of muons, which traverse the calorimeters and the magnet.

The name "Compact Muon Solenoid" refers to the detector's solenoidal magnetic field configuration and its excellent muon tracking capabilities while being more compact (heavier, but smaller) compared to the ATLAS detector.

A cross-section image including the individual subsystems and how different particle types are detected is shown in Figure 3.3. The following sections will discuss the commonly used coordinate system, the design and performance of the detector subsystems as well as the reconstruction of events and physical objects in more detail.

Coordinate Conventions

The CMS Collaboration uses a coordinate system which originates at the nominal interaction point in the centre of the detector. For Cartesian coordinates, the three axes are defined as follows:

- The **x-axis** is horizontal and points roughly towards the centre of the LHC ring.
- The **y-axis** is vertical and points towards the surface.
- The **z-axis** points along the beam axis in westward direction.

Corresponding to the cylindrical shape of the CMS detector, a cylindrical coordinate system is more commonly used. The azimuthal angle ϕ, starting at the x-axis, is used for orientation in the x-y-plane, whereas θ denotes the polar angle starting from the z-axis. The coordinate conventions are visualized in Figure 3.4.

Based on these definitions, the most common coordinate system in CMS physics analyses is defined by the z-axis, the azimuthal angle ϕ and the pseudorapidity η.

Rapidity and Pseudorapidity

The *rapidity* y of a particle is defined as

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) . \tag{3.4}$$

Using rapidity to kinematically describe a particle is often convenient: The rapidity-difference between two objects is invariant under Lorentz-transformations along the z-axis. In addition, at hadron colliders the particle flux is nearly constant as a function of rapidity.

For a particle originating from a fusion of two partons from the colliding protons, the rapidity is given as

$$y = \frac{1}{2} \ln \left(\frac{x_1}{x_2} \right) \tag{3.5}$$

which means that for high absolute rapidity, either $x_1 \ll x_2$ or $x_1 \gg x_2$.

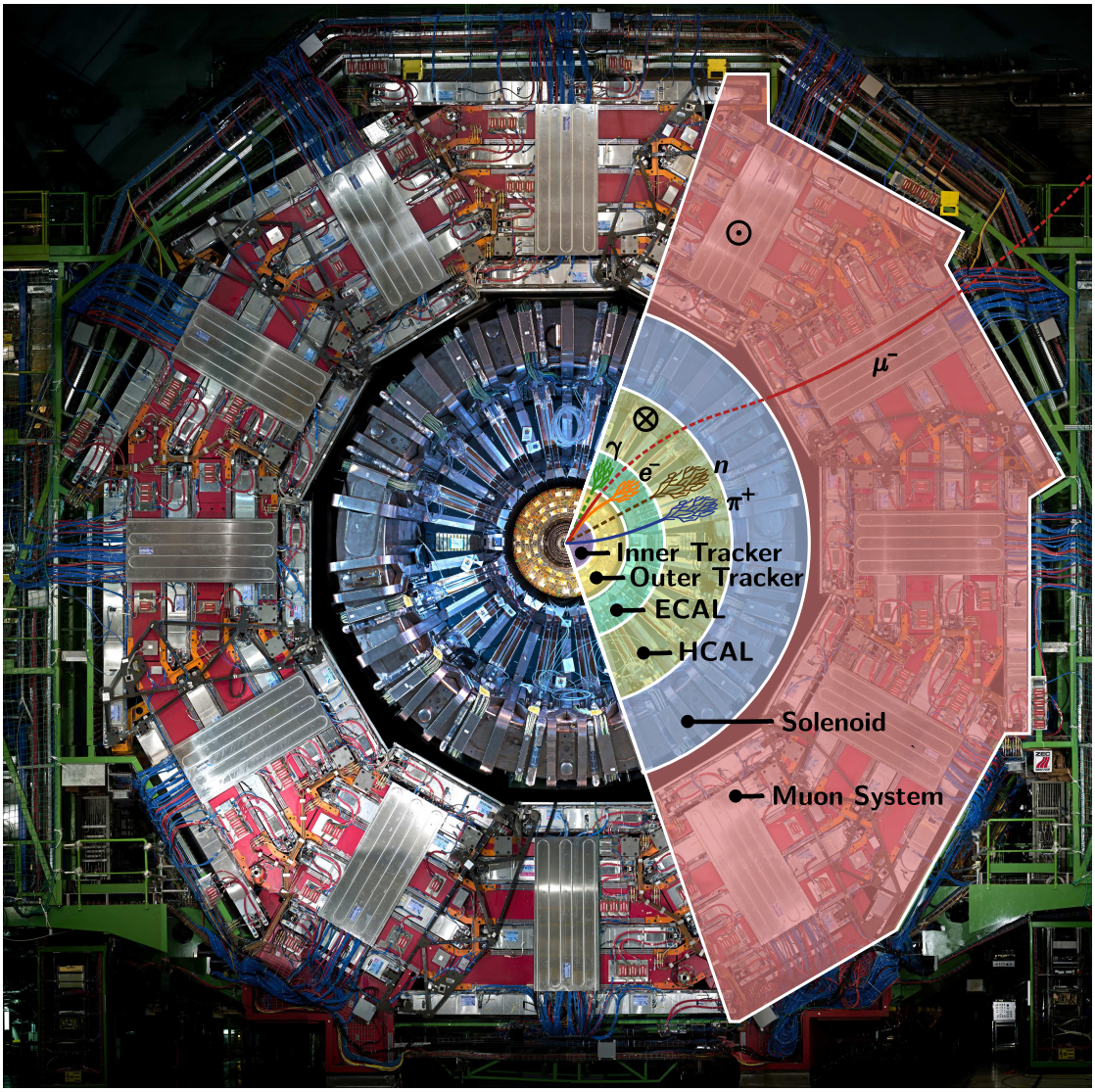


Figure 3.3: A cross section through the central (*barrel*) section of the CMS detector. The different components (inner and outer tracker, electromagnetic and hadronic calorimeters, the solenoid magnet and the muon chambers) are highlighted in different colours. The trajectories of different types of particles are drawn. Solid lines indicate that a particle is detected by the respective detector subsystem, dashed lines indicate that it passes the subsystem unnoticed. The direction of the magnetic field lines is shown by ⊗ and ⊙ inside and outside the solenoid, respectively [43].

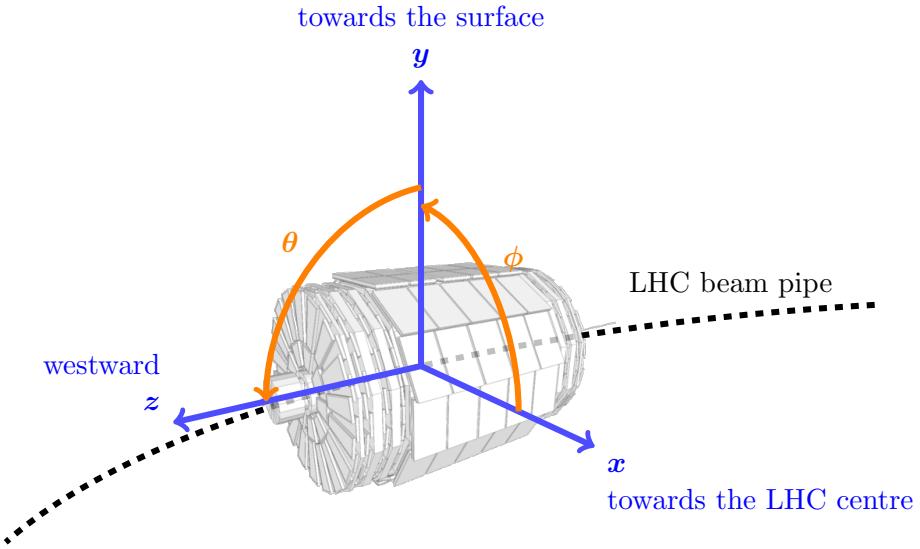


Figure 3.4: Coordinate conventions at the CMS Detector. The x, y and z axes, originating in the centre of the detector, are shown along with the angles θ and ϕ. The orientation is given by the beam pipe, the centre of the LHC ring and the Jura mountains in westward direction. CMS detector sketch from [44].

The *pseudorapidity* η of a particle is defined as

$$\eta = \frac{1}{2} \ln \left(\frac{|\vec{p}| + p_z}{|\vec{p}| - p_z} \right) . \tag{3.6}$$

From this equations 3.6 and 3.5 it can be seen that rapidity and pseudorapidity converge for massless or high-momentum particles:

$$\lim_{p \to E} \eta = y \, . \tag{3.7}$$

Pseudorapidity can also be written as

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) \tag{3.8}$$

which shows that pseudorapidity is a purely geometrical quantity as it only depends on θ. The advantage of using η instead of θ is the more evenly distributed particle flux as a function of η. The relation between pseudorapidity and θ is visualized in Figure 3.5.

The spatial distance ΔR between two objects, e.g. two particles, is commonly measured in η-ϕ-space:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \, . \tag{3.9}$$

In proton-proton collisions, the initial transverse momentum (i.e. transverse to the beam axis, in the x-y-plane) of the interacting partons is negligible. The longitudinal

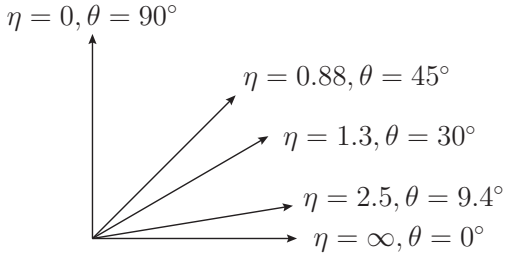


Figure 3.5: Relation between pseudorapidity η and polar angle θ. In the direction of the beam, the pseudorapidity assumes an infinite value [45].

momentum of the initial state is not known since the partons carry an unknown fraction of the proton momentum. Therefore, it is often convenient to kinematically describe a particle produced in a collision in terms of p_{T}, η, ϕ and invariant mass.

3.2.1 Detector Subsystems

Tracker

The task of the tracking system is to measure the trajectories of charged particles and thereby determine their momenta with high precision. In addition, as vertices are reconstructed from tracks, a high spatial resolution is important to distinguish particles originating from the hard scattering, from pileup vertices and from secondary decay vertices.

The CMS tracking system needs to be radiation-hard to withstand the high particle flux close to the interaction point. It should be lightweight to prevent the passing particles from losing energy. At the same time, a high sensor density is required to achieve good spatial resolution. Material and production costs also have to be taken into account. As the best compromise between these requirements, a full silicon-based tracker was chosen.

An outgoing charged particle ionizes the silicon atoms of the tracker, creating charge currents which are amplified and detected. From the positions of these *hits*, the complete track of the particle can be reconstructed (see Section 3.2.2).

The tracking system is 5.8 m long, measures 2.4 m in diameter and extends up to 2.4 in pseudorapidity, making it the largest silicon detector ever built [33]. It is shown schematically in Figure 3.6. The tracker consists of two parts: the inner pixel tracker, close to the interaction region, and the surrounding larger strip tracker.

The **pixel tracker** consists of 66 million pixels, covering an area of 1 m^2. It is located 4 cm to 10 cm away from the beam axis, composed of three central (*barrel*) layers and two endcap disks on either side. It achieves an excellent spatial resolution of 10 µm in r-ϕ-direction and 20 µm in z-direction.

The **strip tracker** consists of silicon strip layers. Because of the lower particle flux further away from the interaction point, the requirement for granularity is not quite as high. Thus, the strip tracker could be constructed more cost-efficient. Still, it achieves a spatial resolution of 23 µm to 52 µm in r-ϕ-direction and 230 µm to 530 µm in z-direction.

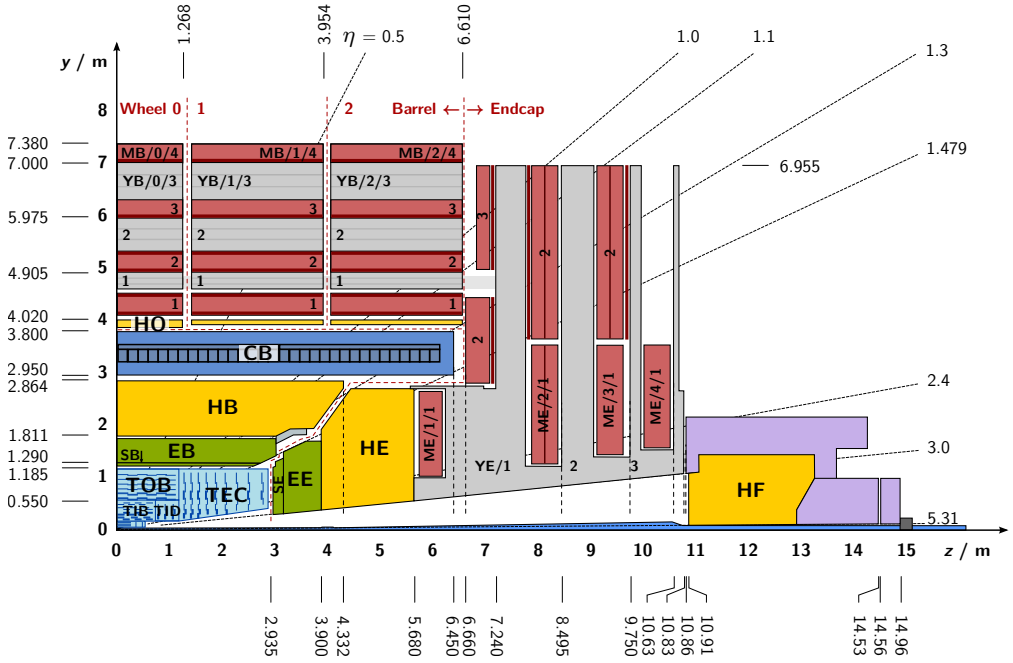


Figure 3.6: Cross section of one quadrant of the CMS detector in the y-z-plane. The different detector subsystems and their size are indicated: The tracking system (light blue), the electromagnetic calorimeter (green), the hadronic calorimeter (yellow), the superconducting solenoid (dark blue) and the muon chambers (red) [43].

Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) measures the energies of photons and electrons. It is the second innermost subdetector, located just outside the tracking system and only 1.3 m away from the interaction point.

Electrons and photons create an electromagnetic shower, effects like pair production and bremsstrahlung lead to a cascade of photons and electrons. Their energy is deposited in the calorimeter material via Compton scattering or the photoelectric effect. The ECAL material scintillates, the resulting photons are caught by avalanche photodiodes and converted into digital signals. The particle energy can be estimated as it is proportional to the number of registered scintillation photons.

The ECAL is made of lead tungstate ($PbWO_4$) crystals because of several reasons: This material has a low scintillation time as 80% of photons are emitted within 25 ns, half the time between two bunch crossings. Also, the short radiation length of 0.89 cm allows for a compact design, with the ECAL fitting inside the magnet coil.

75 848 crystals are arranged in two components, the barrel (EB) and the endcap (EE) sections. These sections cover the rapidity ranges $0 < |\eta| < 1.479$ and $1.553 < |\eta| < 3.0$, respectively. A preshower detector is installed in front of the endcaps to distinguish isolated photons from pion decays. The ECAL can be seen in Figure 3.6, coloured in green, between the tracking system and the hadronic calorimeter.

ECAL calibration is performed with a laser monitoring system and by studying the position of the Z resonance peak in $Z \to e^+e^-$ events. The relative energy resolution of a calorimeter can generally be parametrized as

$$\left(\frac{\sigma_E}{E}\right)^2 = \frac{N^2}{E^2} + \frac{S^2}{E} + C^2 \,. \tag{3.10}$$

N, S and C are terms for noise, stochastic and constant contributions. Equation 3.10 shows that the relative energy resolution of the ECAL improves with increasing energy. An excellent resolution of 1% to 2% in the EB and 2% to 5% in the EE is achieved [46].

Hadronic Calorimeter

The hadronic calorimeter (HCAL) is tasked with measuring the energies of hadrons.

The HCAL was chosen to be a sampling calorimeter: Two alternating material layers act as absorber and scintillator. Therefore, each material can be chosen to best suit its task of high absorption and fast scintillation, respectively. While the high-performance scintillator is made of special plastic, brass has been chosen as the sampling material because of its density resulting in a short interaction length of 16 cm. This enables a relatively compact design, given that hadronic showers typically develop slower than electromagnetic ones.

The HCAL is comprised of several sections to cover as much phase space as possible: a barrel (HB), endcap (HE) and forward (HF) part. An additional outer HCAL (HO) component is located outside the magnet coil. The different components of the HCAL are visualized in yellow in Figure 3.6.

Granularity is highest in the barrel part, achieving a spatial resolution of $\Delta\eta \times \Delta\phi$ = 0.087× 0.087 cm. The relative energy resolution $\Delta E/E$ increases with E and ranges from around 30% to less than 10% [47].

Magnet Coil

The magnetic field is necessary to bend the paths of charged particles. By measuring the curvature of the track, the momenta of particles can be calculated. Because of the high energies of the outgoing particles, a strong magnet field with a large bending force is necessary.

The CMS magnet can produce a magnetic field with a strength of 4 T. It is made out of NbTi and operated at a temperature of 4 K, below the threshold at which the material becomes superconductive. An iron yoke outside of the coil, interleaved with the chambers of the muon system, returns the magnetic field.

Muon System

The outstanding performance in muon reconstruction gives the CMS detector its name. Muons barely interact with the calorimeters or the magnet, their lifetime is sufficiently long to traverse the entire detector before they decay. They can only be identified by their interaction with the muon system.

The muon chambers consist of a barrel and endcap section and cover the pseudorapidity range up to $|\eta| < 2.4$. The different parts of the muon chambers are visualized in red in Figure 3.6.

Because of the large area the muon system has to cover, cost-efficient gaseous detectors have been chosen. When a muon traverses the gas volume, it ionizes the gas atoms. The freed electrons are pulled to an anode, creating a charge pulse in the chamber which is registered as a hit.

The muon system is comprised of three types: resistive plate chambers (RPCs), cathode strip chambers (CSCs) and drift tubes (DTs). DTs and CSCs offer superior spatial resolution, while the strength of the RPCs lies in their fast readout time which is also used for triggering.

For all three types, the time resolution is around 3 ns. The spatial resolution in the DTs is 80 µm to 120 µm, 40 µm to 150 µm in the CSCs and 0.8 cm to 1.2 cm in the RPCs [48].

Event Trigger

The data recorded by all subsystems of the CMS detector for each bunch crossing amounts to roughly 1.5 MB. If collisions take place every 50 ns, 30 TB of data are produced per second – an unmanageable amount.

However, only few events are of scientific interest (which is what necessitates the high collision frequency in the first place) as only collinear or elastic scattering takes place during most bunch crossings.

The trigger system is tasked with selecting the events which are potentially relevant for physics analyses. These decisions have to be made very quickly. Therefore, a combination of hardware and software triggers is used.

The Level-1 (L1) hardware trigger is fast and reduces the event rate to around 100 kHz. The software-based High Level Trigger (HLT) system uses more complex algorithms, which can be updated and improved, and further reduces the event rate to roughly 150 Hz. As a result of selecting only interesting physics events, 225 MB/s are selected for permanent storage.

3.2.2 Reconstruction of Physical Objects

The output of the detector is a bunch of electronic signals from the various subsystems. At a basic level, the detector data contains information on hits in the tracking system or energy deposits in the calorimeters. For physics analyses, physical objects, e.g. particles or jets, have to be reconstructed from this data.

In the following sections, it is described how the CMS Collaboration uses the Particle-Flow technique for event reconstruction. The influence of pileup interactions also has to be taken into account.

The Particle-Flow Method

For the reconstruction of physical objects, the Particle-Flow method [49, 50] is used by the CMS Collaboration. It aims to exploit and combine as much information from the different detector subsystems as possible.

The event reconstruction starts with muons, combining hits in the muon chambers and the tracker. From energy deposits in the calorimeters and tracks, charged and neutral particles are reconstructed. The reconstructed muons, electrons, photons, charged and neutral hadrons are referred to as *Particle-Flow candidates*. Jets are reconstructed from the collection of Particle-Flow candidates. This is in contrast to previous experiments or other analysis techniques where jets are reconstructed from calorimeter clusters. With the Particle-Flow method, the event and object reconstruction performance was improved compared to other techniques [51].

Particle Track Reconstruction

The CMS tracker, described in Section 3.2.1, detects only individual *hits* of a particle traversing the tracker material. The reconstruction of complete particle tracks from the collection of hits is a combinatorial problem which is solved by a *track-finding algorithm*. The CMS Collaboration uses the Kalman filtering technique [52, 53] for this purpose.

The track reconstruction start with clusters of three hits in adjacent detector layers, so-called triplets, in the high-resolution pixel detector. These are used as *seeds*, starting points for the reconstruction procedure of the entire track. From the positions of a triplet's hits, an initial track direction and curvature can be determined. The track is extrapolated to the outward or inward detector layers. If matching hits are found, they are added to the track.

The efficiency for track reconstruction is high, amounting to 94% in the central detector region at $|\eta| < 0.9$ and 85% at $0.9 < |\eta| < 2.4$. The p_{T} resolution worsens at higher absolute pseudorapidity and at higher momenta due to the lower track curvature and is in the range of 0.6% to 10% [54].

The magnetic field forces charged particles on a curved trajectory via the Lorentz force, which depends on the (known) magnetic field strength and the momentum of the particle. Therefore, the particle momentum can be derived from the curvature of the track. Combining this information with the particle energy known from matching calorimeter deposits, the kinematic quantities of the particle can be fully determined.

Vertex Reconstruction

The primary interaction vertices are reconstructed from the collection of reconstructed tracks. Tracks which originate from a similar position in the interaction region are clustered and the vertex position is fitted.

The spatial resolution of the vertex position improves on the number of tracks associated with the vertex. For hard-scattering vertices associated with at least 50 tracks, it can be as good as 10 µm in x and y and 12 µm in z direction [54]. The signal vertex is defined

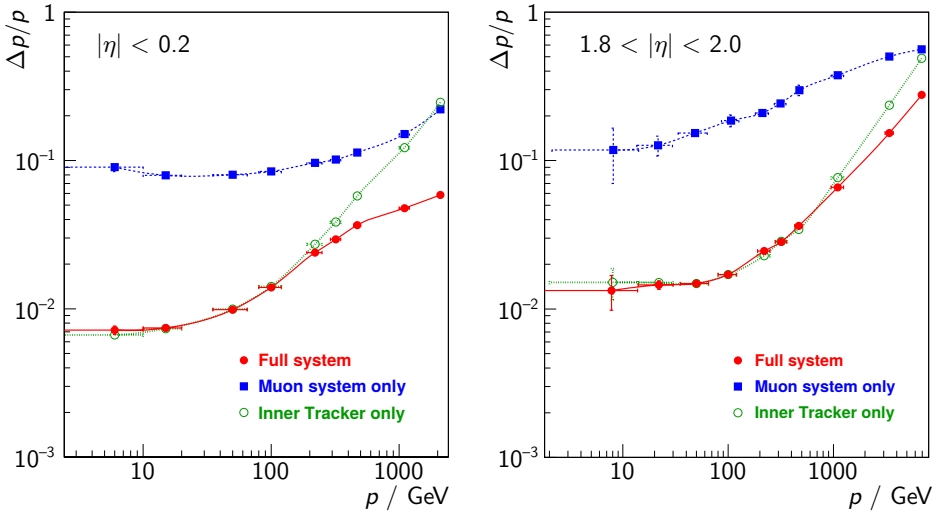


Figure 3.7: Muon p_T resolution for muons reconstructed in the barrel (left) and endcap (right) region. The best resolution is achieved when information from the muon chambers and the tracker is combined [56].

as the vertex with the highest $\sum p_T^2$ of the associated tracks, as this is most likely the vertex of the hard interaction.

Muon Reconstruction

Muons are reconstructed by combining reconstructed tracks in the tracker and the muon system. Because muons are reconstructed from two complementary detector systems over a large distance, high reconstruction efficiency and a precise measurement is ensured.

Muon trajectories are first reconstructed in the muon chambers, using a similar seeding method as for the tracker. The reconstructed track is extrapolated from the muon chambers to the inner tracking system and associated with to matching hits. Effects such as energy loss due to bremsstrahlung, multiple scattering in the detector or inhomogeneities in the magnetic field are taken into account by the reconstruction algorithm.

Due to the low activity in the muon chambers, the muon reconstruction efficiency is high while the fake rate is low. Muons are usually subjected to a series of quality criteria to be eligible for physics analyses, detailed in Appendix A.1.

The resolution of the muon momentum worsens at higher muon energies: with increasing transverse momentum of the muon, the curvature of the track becomes smaller. As the curvature approaches the spatial resolution limit of the track reconstruction, the momentum determination becomes less precise. The relative muon momentum resolution is shown in Figure 3.7. For most of the phase space it is at the level of only few percent [55].

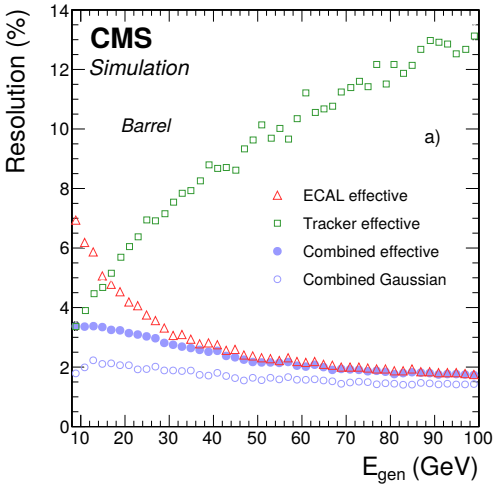


Figure 3.8: Electron relative p_T resolution as a function of the simulated electron energy. The resolution from the track worsens at higher momenta, while the calorimetric resolution improves. The total resolution is determined by combining the ECAL and tracker measurements. Over most of the phase space, the superior calorimetric resolution drives the combined resolution [57].

Electron Reconstruction

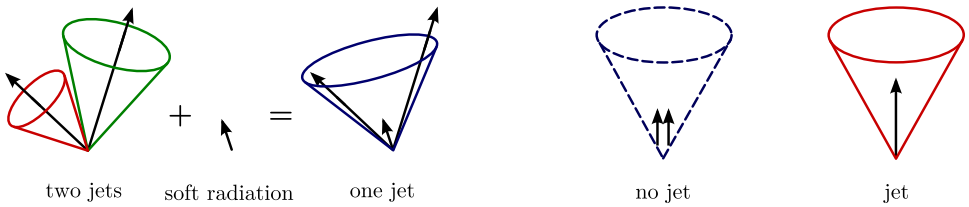
Electrons are reconstructed by combining tracks and ECAL energy deposits. The performance of electron reconstruction and calibration is detailed in [57].

The reconstruction of electrons is complicated by the large energy losses because of bremsstrahlung. On average, 33% of the electron energy are radiated where material is minimal ($|\eta| = 0$) and 86% where it is thickest ($|\eta| = 1.4$).

The reconstruction in the ECAL starts with a seed crystal, an ECAL crystal with a high energy deposit. Neighbouring clusters are added if their energy exceeds a threshold, forming a *supercluster*. The η and ϕ coordinates of the supercluster are determined from the weighted mean of the energy distribution.

For the electron track reconstruction, the Kalman filter cannot be used because of the large radiation. Instead, a *Gaussian Sum Filter* [58] is used which takes radiative energy losses into account. The seed of the track is extrapolated from the coordinates of the ECAL supercluster or from the first hits in the tracker. Finally, the track and supercluster are combined via geometrical matching.

The relative p_T resolution of electrons is visualized in Figure 3.8. For electrons with p_T below 15 GeV, the momentum determination is mainly track-driven. At higher p_T, calorimetric resolution increases and the ECAL energy measurement provides a preciser estimation of the momentum.



(a) If soft radiation affects jet clustering, the used algorithm is infrared unsafe.

(b) If collinear splitting affects jet clustering, the used algorithm is collinear unsafe.

Figure 3.9: Collinear and infrared unsafety for jet clustering [59].

Jet Reconstruction

Partons from a hard interaction fragment and hadronize into a stream of collimated particles, a jet. As the experimental signatures of quarks and gluons, jets are important objects for many physics analyses.

Concerning reconstruction, there is no unambiguous answer to what a jet is. Multiple definitions exist, depending on the type of input objects and the algorithm used to combine them.

In simulation, jets can be formed by combining the four-vectors of simulated particles (*particle jets*). In data, jets are reconstructed by combining Particle-Flow candidates. These *Particle-Flow jets* have to pass a set of quality criteria to be eligible for physics analyses, see Appendix A.1.

From the list of particles, jets are reconstructed by a *jet clustering algorithm*. Many of these exist, leading to different types of jets. The most important algorithms are discussed in the following section.

Jet Clustering Algorithms

Historically, *fixed cone algorithms* have been used as the most basic category of jet clustering algorithms: All particles within a distance R to a given axis are part of a jet. As an extension, *iterative cone algorithms* determine the jet axis iteratively from the four-vector sum of the clustered particles

However, fixed or iterative cone algorithms are not robust enough to provide collinear or infrared safety, important requirements for comparability between data and predictions from perturbative QCD. *Infrared safety* means that soft radiation in the final state does not influence jet clustering, see Figure 3.9(a). *Collinear safety* means that the occurrence of collinear splitting does not influence the jet clustering, see Figure 3.9(b).

As a refinement of basic cone algorithms, seedless infrared-safe (SIS) cone algorithms have been developed to fulfil both requirements [60].

A slightly different approach while also being infrared and collinear safe is used by the class of *sequential recombination algorithms*: These do not have a fixed cone size, but

use a bottom-up approach, starting from the list of particles and combining them until certain criteria are fulfilled.

The family of generalized k_T algorithms [61] combines particles via the following distance measure:

$$d_{ij} = \min(p_{\mathrm{T},i}^{a}, p_{\mathrm{T},j}^{a}) \frac{\Delta R_{ij}^2}{R^2} \tag{3.11}$$

where R denotes the jet size parameter, and ΔR_{ij} is the spatial distance between two objects i and j. Particles are added to the jet until all distances d between jet and remaining particles exceed a minimum threshold $d_{\mathrm{iB}} = p_{\mathrm{T},i}^{a}$.

a is a parameter to set the influence of the p_{T} scale relative to ΔR. Typically, the following values are used:

- $a = 2$, resulting in soft k_T jets.
- $a = 0$ for p_{T}-independent clustering, known as the Cambridge/Aachen [62, 63] algorithm.
- $a = -2$ for anti-k_T [64] jets, favouring the clustering of high-energetic particles.

The anti-k_T algorithm produces jets with fairly cone-like shapes and similar sizes in η-ϕ-space. This improves the comparability between the individual jets and simplifies tasks like jet corrections.

Anti-k_T jets with a jet size parameter of $R = 0.5$ have in many cases shown to be a reasonable compromise between catching as many particles from the hard interaction as possible and catching as few pileup particles as possible. Therefore, it is the most widely used jet size parameter in the CMS Collaboration. Jets with size parameters of $R = 0.4$ and $R = 0.6$ are commonly used in physics analyses by the ATLAS Collaboration [65].

For some studies, a jet size parameter of $R = 0.7$ is used, as fixed-order perturbative QCD calculations often perform better for larger jet size parameters [66].

Jet Area and Event Energy Density

An important quantity in the context of jet studies is the jet area A_{j} in η-ϕ-space [67]. It is especially of interest for jet energy corrections as the contribution from pileup roughly scales with the area of the jet. As sequential recombination algorithms do not have a fixed jet size, the occupied area is not known a priori but has to be measured.

The jet area is determined by adding a large number of infinitesimally soft particles to the event, evenly distributed in η-ϕ-space. The inclusion of these particles does not significantly change the properties of the jet. However, the number of these particles included in the jet is proportional to the occupied area in η-ϕ-space.

Another important quantity is the average p_{T} density per event, denoted ρ [68]. It is defined as the median of the $p_{\mathrm{T,j}}/A_{\mathrm{j}}$ distribution of jets clustered with the k_T algorithm using a size parameter of $R = 0.6$. This algorithm produces many soft jets, effectively covering the entire η-ϕ-space.

As this includes the energy from the underlying event and from pileup interactions, ρ is a measure for the activity in an event, useful e.g. in the context of jet energy corrections.

Missing Transverse Energy

The missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the negative of the four-vector sum of all reconstructed particles.

In the initial state of the collision, transverse momentum arises only from the small intrinsic parton momenta. Because of momentum conservation, the vector sum of all outgoing particles also should have only a negligible transverse component.

Minor $E_{\mathrm{T}}^{\mathrm{miss}}$ can be measured in most events, caused by detector miscalibration or particles escaping the detector acceptance. However, a significant $E_{\mathrm{T}}^{\mathrm{miss}}$ is usually a sign of undetectable particles such as neutrinos in the final state. Therefore, $E_{\mathrm{T}}^{\mathrm{miss}}$ is especially important for analyses of such processes, e.g. top quark or τ decays.

The performance of $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction and resolution has been extensively studied by the CMS Collaboration, the measured results have been generally found to agree with simulation [69, 70].

Pileup Mitigation

Pileup effects can cause a large systematic bias as they overlay the signature of the hard interaction. The mitigation of pileup is conducted on several levels.

In addition to pileup collisions in the same bunch crossing (in-time pileup), *out-of-time pileup* refers to collisions in the previous or following bunch crossing. These effects are due to pileup affecting adjacent bunch crossings because of the finite calorimeter readout time. On hardware level, the optimization of the calorimeter readout time can decrease the influence of out-of-time pileup.

On software level, Charged Hadron Subtraction (CHS) was the most used tool during the 2012 data-taking period. Charged hadrons which can be traced back to a vertex other than the signal vertex are removed from the list of Particle-Flow candidates prior to jet clustering. Thus, the pileup influence on jets is decreased. The CHS method is visualized in Figure 3.10.

For jet reconstruction, the pileup energy contribution has to be subtracted from the jet, see Section 5.1.1. A multivariate discriminator to identify jets originating from pileup vertices has also been developed by the CMS Collaboration [71].

For Run 2 of the LHC, *Pileup Per Particle Identification* (PUPPI) [72] has been developed to optimize the estimation of pileup energy. Every particle is assigned a weight depending on its probability to originate from a pileup interaction. The event is then reinterpreted, e.g. jets are reclustered while taking these weights into account.

3.2.3 Luminosity Measurement

As the luminosity depends on parameters of the LHC beam (see Equation 3.3) which are not perfectly known, e.g. the exact number of protons per bunch, the luminosity can only be roughly predicted from machine parameters and has to be measured with the detector.

Because the luminosity relates the measured event rate to the cross section, it is an important ingredient for all cross section measurements. For all such analyses, the

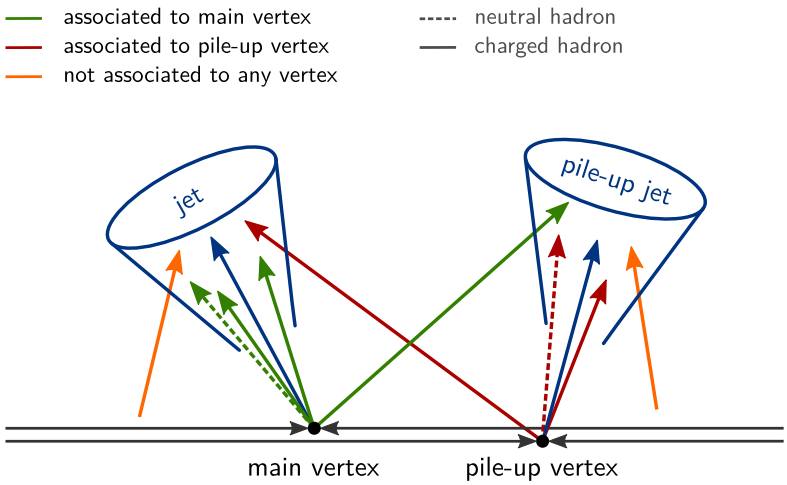


Figure 3.10: Jet clustering and Charged Hadron Subtraction (CHS). The particles which can be traced back to pileup vertices (indicated in red) are removed by CHS prior to jet clustering [43].

uncertainty on the luminosity is an important contribution to the total uncertainty, sometimes even the largest one.

The CMS experiment measures the luminosity by exploiting the proportionality between luminosity and particle flux in the detector [73]. The counting of signal clusters in the inner pixel detector serves as the main estimator. The occupancy in the hadronic forward calorimeters (which can be measured less precisely because of possible time dependence of the calorimeter response) is used as a cross check.

However, these methods only achieve a *relative* measurement of the luminosity for different periods. An *absolute* normalization is performed with *van der Meer scans* [74]: Dedicated LHC runs with different degrees of beam separation are performed. By measuring the interaction levels for different separation degrees and combining the results, the absolute luminosity can be determined. As a result, the luminosity can be measured with an uncertainty of 2.5% (syst.) + 0.5% (stat.) [73].

4 Software Tools for High Energy Physics Analyses

High energy physics relies upon data-intensive computing: The output signals from the detector have to be processed, simulations have to be computed, numerous calculation steps are necessary to obtain the final results of an analysis. To cope with these challenges, a multitude of tools and programs is needed, some of which have to be developed by the analysis groups themselves. In the following sections, an overview of the computing steps and the used software is given.

4.1 Analysis of Detector Data

The LHC provides millions of proton collisions per second, the CMS detector records 1.5 MB of data per event. Even though the trigger system selects only the events potentially relevant for physics analyses, the size of the stored data is still in the order of many petabytes. In addition to these storage demands, the necessary steps to extract physical results out of these data require immense computational resources.

The entire data processing chain involves a multitude of settings. As a physics analysis evolves, many parameters change and additional demands arise, leading to frequent reprocessing. Therefore, the data processing is commonly split into several steps, with intermediate results being stored in dedicated formats. This splitting is done such that typical parameter changes necessitate only the reiteration of the latter steps, thus decreasing turnaround time and overall computing effort.

The data are first processed with the official software framework of the CMS Collaboration. Often, the final processing steps are conducted with tools specifically developed for a particular analysis and written by the analysis group itself. A typical analysis workflow is visualized in Figure 4.1.

4.1.1 The CMS Software and Computing Infrastructure

The data recorded by the CMS detector and simulated samples are stored at large computing centres in different countries, organized into the Worldwide LHC Computing Grid (WLCG) [75]. The size of the data imposes high storage and computing demands on the WLCG.

The data are stored in different formats, from the raw detector output to formats where reconstruction algorithms have been applied and only information on high-level objects (reconstructed particles etc.) is available. The file format of the ROOT [76, 77] framework is used for storage.

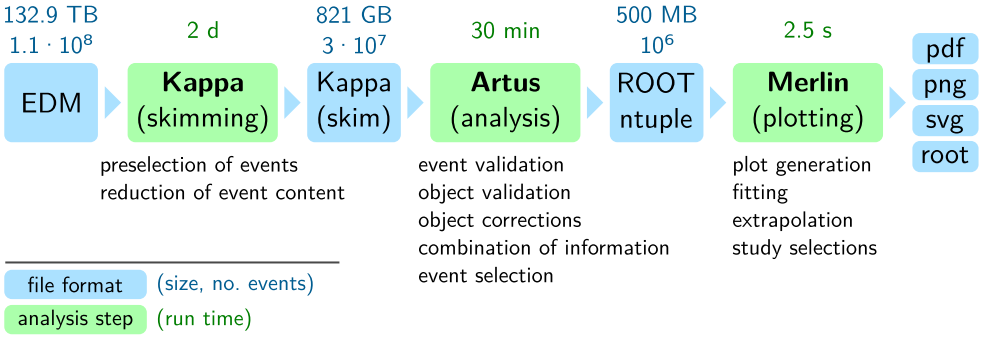


Figure 4.1: Workflow of a physics analysis at the CMS experiment, using the Artus software suite. The processing is performed in several steps, from the detector data (EDM) to the graphical representation of results [43].

The CMS Collaboration uses the concept of the Event Data Model (EDM), where one object contains all the event data and can be handled independently from all other events. The data can be processed with CMSSW [78], the official CMS software framework. CMSSW has a modular structure, with different modules for e.g. reading of the stored detector data, performing reconstruction and event selection steps or storing the output of the processing. In addition to the officially provided modules, analysis groups can add their own modules according to their diverse needs.

CMSSW is also equipped with an interface to various Monte Carlo event generators, simplifying the simulation of physics processes, the subsequent detector simulation and the storage of the results in the EDM format.

Commonly, CMSSW is used to perform a loose event selection and store the results in a dedicated format for further processing, a step referred to as *skimming*.

4.1.2 Analysis with Kappa, Artus and Excalibur

For the remaining processing steps, the analyses detailed in this thesis utilize the Kappa [79], Artus [80] and Excalibur [81] frameworks.

Kappa is a software package for skimming. Handling the output of the CMSSW processing, only analysis-relevant information on high-level objects are stored in Kappa's own data format. The selection of object types for storage is configurable, thus Kappa can be flexibly used for a wide range of physics analyses.

Kappa is designed to reduce disk space requirements: Typically, the size of the skimming output is several orders of magnitude smaller compared to the original data. Additionally, as the speed of the data processing is typically limited by input-output operations, the reduction in size greatly increases the speed of further processing steps. With the improvements in convenience, storage needs and turnaround time, Kappa lays the foundations for efficient analysis work.

Further processing is performed with Artus, a state-of-the-art framework for data analysis in high energy physics. It handles all remaining aspects regarding event reconstruction and event selection. Artus is mainly written in C++ and has been collaboratively developed at KIT according to the latest and highest standards of software design and engineering.

Typically, software for a particle physics analysis is continuously written alongside the analysis effort, without an existing plan or design concept. This approach naturally leads to code which is inflexible, unstructured and difficult to maintain. Artus avoids these problems and surpasses previous analysis tools in terms of performance, maintainability and reliability. Due to its sophisticated structure and the pipeline concept, turnaround time is reduced to a minimum. A clear, object-oriented concept and the use of software design patterns leads to uncluttered and well-structured code. A more detailed overview of its architecture and the underlying design principles is given in [80].

Artus is used by several analysis groups with a large and active user base, its key components are therefore well tested. As the various analysis steps are separated into dedicated classes and new modules can easily be added, Artus is conveniently extendable by new or existing analysis groups. An extensive documentation including tutorials and working examples provides a rich and smooth learning experience for newcomers. Artus has been thoroughly tested and successfully used in a number of physics analyses, e.g. Higgs boson or jet measurements [82, 83].

As Artus follows the EDM, i.e. events are handled independently of each other, the event processing can be trivially parallelized by splitting the input data set. This can be easily done with grid-control [84] which is natively supported by Artus. For the analyses presented in this thesis, Artus has been executed in parallel and utilizing a dedicated caching system [85, 86] for even higher performance.

For the graphical representation of analysis results, the HarryPlotter program can be used. It is part of the Artus suite and optimized for working with Artus output files. It follows a similar design concept and is highly extendable and configurable. It can read data from various sources (e.g. ROOT or YODA [87] files), perform additional analysis steps (e.g. histogram normalization or ratio calculation) and produce output graphics in different formats (e.g. PDF or PNG files via matplotlib [88]).

The Excalibur package is an extension for Artus, adding functionalities necessary for Z+jet studies such as jet calibration or selection of Z+jet events. Additional scripts and tools, e.g. for configuration, are also included.

Merlin is an add-on for HarryPlotter, contained in the Excalibur package. It extends HarryPlotter by adding helpful features for graphical representation in the context of Z+jet studies. It was used for almost all graphs relating to analysis results in this thesis.

4.2 Simulation of Physics Processes

Based on the theoretical knowledge on particle physics, it is possible to simulate physics processes like proton-proton collisions. This includes the calculation of observables such as cross sections for particular processes, enabling comparisons between theory and experiment.

As these calculations can become arbitrarily complex and involve multi-dimensional integrals, they can usually not be solved with analytic methods. Instead, a numerical integration is performed via Monte Carlo techniques.

The computational tools for these methods are Monte Carlo simulation programs, also referred to as event generators. Using perturbative QCD, they are able to describe a number of processes at leading- or next-to-leading-order accuracy. They can be specialized in different stages of the event generation process, e.g. matrix element calculation or parton showering.

General information on Monte Carlo event generators and their usage can be found in [89]. Examples for Monte Carlo programs are Sherpa [90], Pythia [91], Herwig [92], MadGraph [93] or FEWZ [94].

To compare data and simulation, the results of the theory calculation (i.e. four-vectors of stable particles) are often subjected to a simulation of the detector. This simulation is performed with a program such as Geant4 [95] which calculates the interactions of particles with matter. As a detailed model of the geometry of the CMS detector is available, the output of the detector simulation can be compared to data.

The large-scale generation of events, especially in combination with the subsequent detector simulation, requires large computational resources, rivalling the analysis of detector data in terms of storage and computing needs. The simulation of $Z(\rightarrow e^+e^-)$+jet events with Sherpa for the studies detailed in Chapter 6 has been performed in parallel on high-performance workstations organized into a computing cloud at KIT [96]. The corresponding Sherpa configuration file is shown in Appendix A.3.1.

4.3 Tools for PDF Studies

PDFs can be determined by comparing data with theory predictions. For fast cross section predictions based on different PDFs, the calculations have to be available in dedicated formats such as fastNLO interpolation tables. The PDF determination can be performed with the xFitter framework. The necessary steps can be simplified with the SheRivF toolkit [97]. A typical workflow for PDF studies as used in Chapter 6 is shown in Figure 4.2.

fastNLO, Rivet and MCGrid

Fits of PDFs to experimental data require theory predictions based on different PDFs. Full calculations (especially beyond leading-order accuracy) of an observable distribution for a large number of PDFs are far too time-consuming. Instead, theory predictions for different PDFs or at different scales can be quickly created with interpolation techniques,

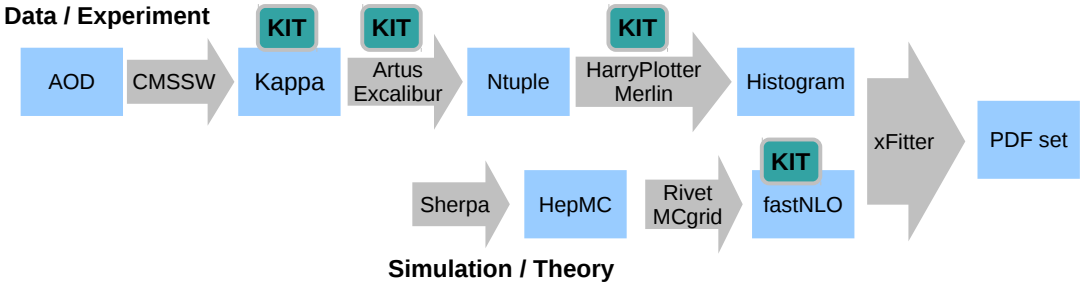


Figure 4.2: $Z(\rightarrow e^+e^-)$+jet analysis workflow. Several steps for data analysis and simulation are performed. xFitter is used to combine measurement and theory predictions to determine the PDFs. Many of the used tools have been (co-)developed at KIT.

e.g. using the fastNLO [98] software. In principle, the cross section calculation is split into constant and flexible parts. In QCD factorization, the cross section can be written as an integral of the PDFs over x, see Equation 2.8. fastNLO replaces each PDF with a sum of simpler interpolation functions. Thereby, the cross section can also be written as a sum, depending on PDFs, scales, α_s (which are all flexible) and constant coefficients which are stored in a look-up table. As a result, the calculation of the cross section is reduced to evaluating this sum by specifying the PDFs, scale and α_s values; a procedure which is much faster than integration. An easy interface to different PDF sets is provided by the LHAPDF [99] tool.

The fastNLO coefficient tables can be created from simulated events using Rivet. Rivet is a validation tool which can be interfaced to various Monte Carlo simulation programs. The generated events are passed to Rivet where additional analysis steps, e.g. phase space restrictions, are applied to achieve comparability with experimental results. The fastNLO tables are then created via the Rivet-plugin MCGrid [100, 101].

xFitter

xFitter [102, 103] (formerly HERAFitter) is a framework to determine PDFs by comparing data and theory. With its modular design, it is capable of different approaches for PDF determination and can handle theory predictions in different formats.

In a typical usage scenario, PDFs are parametrized at a starting scale and evolved to the scale of the measurement via the DGLAP equations in the $\overline{\text{MS}}$ scheme [104], implemented via the QCDNUM [105] package. QCDNUM is also used for calculating the structure functions and thereby the cross sections for deep-inelastic scattering processes. The PDF parameter values are then fitted using Minuit [106] to minimize the specified goodness-of-fit estimator for the agreement between data and theory.

5 Calibration of the Jet Energy Scale with $Z(\rightarrow \mu^+\mu^-)$+Jet Events

A high-energetic hadron collision can lead to the emergence of isolated partons. These partons fragment and hadronize into jets, streams of collimated particles. Jets are important experimental signatures: Almost all LHC physics analyses rely on jets for signal extraction or background suppression. Because of the large number of pileup interactions likely to produce isolated partons, jets are ubiquitous in collisions at the LHC.

Unfortunately, the measurement of jets is biased by many systematic effects: Particles from the hard scatter might falsely not be part of the reconstructed jet (out-of-cone effects). Particles from pileup interactions or the underlying event might mistakenly be clustered into the jet, increasing the measured jet energy, Energy might be lost through initial and final state radiation Nonlinear calorimeter response or electronic noise in the detector also lead to systematic biases.

These effects lead to large systematic uncertainties on the jet energy scale (JES). For physics analyses that rely on jets, the resulting uncertainty on a measured quantity, e.g. a cross section, can be even larger. For many measurements at the LHC, the JES uncertainty is the dominating experimental uncertainty.

To reduce this uncertainty, the JES has to be carefully calibrated. The calibration procedure is a complex task: Because of the multitude of effects that bias the jet measurement, several sophisticated techniques have to be used to estimate the contributions from the different uncertainty sources. Correction factors are derived by combining the results from different analyses. The procedure of jet reconstruction and jet energy correction is visualized in Figure 5.1.

For a reliable result, the use of data-driven calibration methods is indispensable. One of these methods utilizes events with a Z boson produced in association with a jet. The Z boson decay into a pair of muons or electrons leaves a clear signature in the detector. Because of the excellent resolution in electron or muon measurement, the Z boson can be reconstructed with high precision. The transverse momenta of jet and Z boson in Z+jet events are related because of momentum conservation. This relation can be exploited to precisely estimate the jet energy and it is therefore used for calibration.

The following sections will outline the jet calibration procedure carried out by the CMS Collaboration, the $Z(\rightarrow \mu^+\mu^-)$+jet analysis and the data-driven determination of the absolute JES.

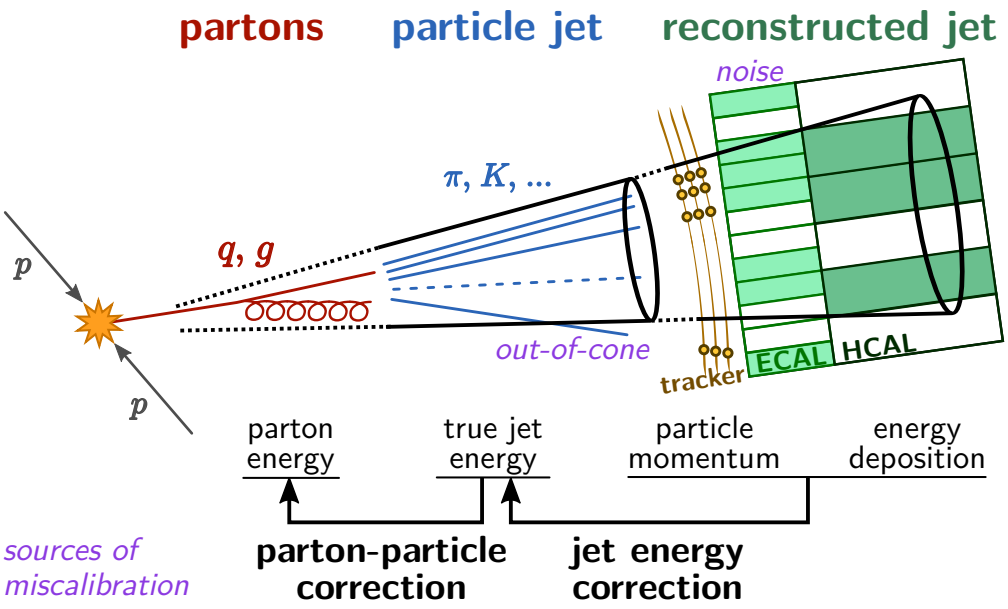


Figure 5.1: Jet reconstruction and correction. A parton fragments and hadronizes into stable particles. Momenta and energies of jet particles are measured in the tracking system and in the calorimeters, respectively. Corrections of the jet energy relate the reconstructed jet momentum to the particle-level momentum ("true jet energy"). An additional parton-particle correction can further correct to the level of parton energy [43].

5.1 Jet Energy Calibration for the CMS Experiment

The reconstruction of jets with the CMS detector has already been outlined in Section 3.2.2. This section will explain the calibration of jets in the proton-proton collision data recorded by the CMS experiment at $\sqrt{s} = 8\,\text{TeV}$, corresponding to an integrated luminosity of $19.7\,\text{fb}^{-1}$.

The correction of the jet energy is performed in several successive steps, see Figure 5.2. Each step addresses a dedicated problem:

1. **Pileup offset correction** Subtraction of the additional energy from particles produced in pileup interactions.

2. **Simulation-based correction** Correction for reconstruction biases that are known and can be simulated.

3. **Data-driven correction** The correction from simulation has to be complemented with data-driven methods.

4. **Flavour correction** Correction for the JES differences between jets originating from different parton flavours.

Steps 1-3 are mandatory for all CMS analyses involving jet measurements. Step 4 can be optionally applied: Usually, flavour corrections are only needed for analyses with a

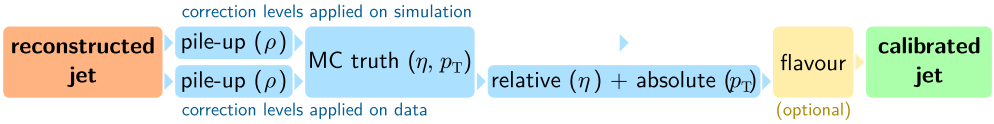


Figure 5.2: Jet energy correction steps in the CMS Collaboration. Reconstructed jets in data and simulation are corrected for pileup offset and then scaled to particle("MC truth")-level. Additional corrections for the relative and absolute JES are applied to data. Optionally, flavour corrections can be applied. Only corrected jets are eligible for physics analyses [43].

focus on particular jet flavours, e.g. studies on vector boson production in association with jets from heavy quarks [107].

The advantage of this factorized procedure is that dedicated methods can be used to derive the correction for each step. Also, the calibration for the latter steps can cross-check the correction from the former steps, ensuring high reliability and precision.

The entire calibration effort comprises the work of several analysis groups within the CMS Collaboration. The methods and results from the calibration of the data recorded at $\sqrt{s} = 7\,\text{TeV}$ and $8\,\text{TeV}$ are detailed in [108] and [109], respectively. Intermediate performance reports for the $8\,\text{TeV}$ data are available at [110–114]. An overview of the jet calibration results from the ATLAS and CMS Collaborations is given in [115].

In the following sections, the procedures to derive the correction for pileup offset, the simulation correction and the data-driven correction are outlined.

5.1.1 Corrections for Pileup Offset

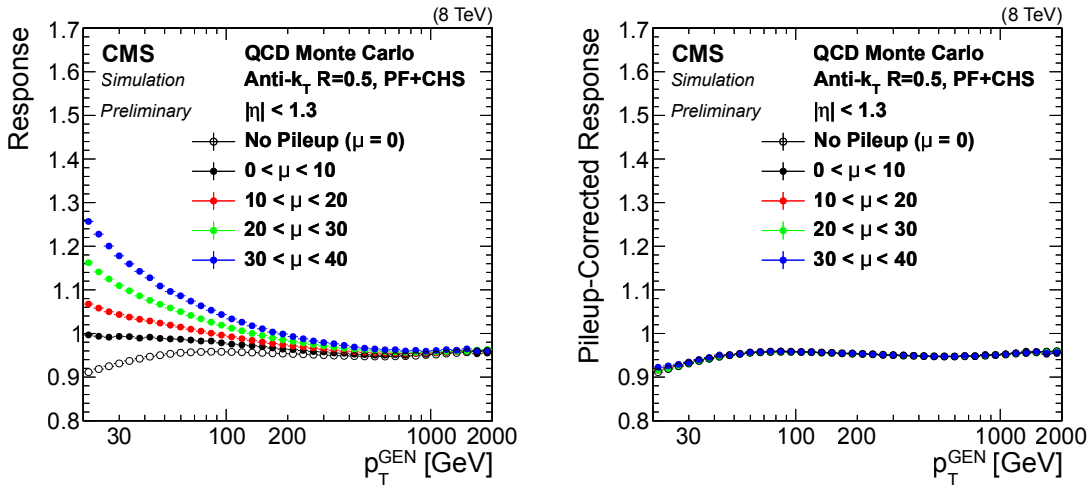
In the data at $\sqrt{s} = 8\,\text{TeV}$, on average 20 pileup collisions take place in each bunch crossing (see Section 3.1) leading to a dense background of pileup particles. Even though charged hadrons originating from pileup vertices are ignored during jet clustering (see Section 3.2.2), a significant amount of pileup energy contributes to the measured jet energy. Therefore, the energy offset from pileup has to be subtracted as the first correction step.

The challenge of this correction level is the determination of the pileup energy contribution for a given jet. The *hybrid jet area method* achieves a precise estimation of the pileup offset by taking the effective jet area A_j and the average energy density per event ρ (see Section 3.2.2) into account. The method is named "hybrid" because it combines the original, η-independent approach [116] of only using A_j and ρ with the η-dependence necessary to account for detector geometry:

$$c_\text{hybrid}(A_\text{j}, \rho, \eta, p_\text{T}^\text{raw}) = 1 - \frac{[\rho_0(\eta) + \rho \cdot \beta(\eta) \cdot (1 + \gamma(\eta) \cdot \log(p_\text{T}^\text{raw}))]}{p_\text{T}^\text{raw}} \tag{5.1}$$

where β, ρ_0 and γ are parameters needed to adjust the shape of the correction to the observed dependence in η. The contribution from the underlying event is absorbed by these parameters and their values into the correction factor c_hybrid.

The values for the parameters are determined from the particle-level offset in simulation. For data, additional scale factors are derived using the *Random Cone* method: A sample



(a) Before pileup corrections

(b) After pileup corrections

Figure 5.3: Jet response in simulation as a function of jet p_T for different pileup scenarios. After applying corrections for pileup, the jet response for the different pileup scenarios agrees [114].

of proton-proton collision data with only soft proton interactions and no hard scattering process is used. Jets are reconstructed in randomly placed cones until the entire η-ϕ-space is covered. As there is no hard scattering in these events, the average p_T of these jets corresponds to the average pileup offset in data.

Figure 5.3 shows the simulated response for central ($|\eta| < 1.3$) jets in different pileup scenarios, parametrized via the number of pileup interactions per bunch crossing μ. For low-energetic jets, the effect from pileup offset is particularly large. After the correction for pileup, the response for the different pileup scenarios agrees, demonstrating the success of the correction in reducing pileup dependence.

While the response is now largely independent from pileup, a significant p_T dependence, with the jet response systematically below one, remains. This clearly shows the need for further correction steps.

5.1.2 Corrections Based on Simulation

After subtracting the offset energy from pileup, simulation-based jet corrections are applied. They cover the bulk of the systematic effects and leave only a small residual correction for the data-driven calibration methods. The advantages of the simulation-based corrections are that they rely on additional information from simulation, they are not sensitive to many of the biases of data-driven methods and they can cover phase space regions which are not easily accessible in data.

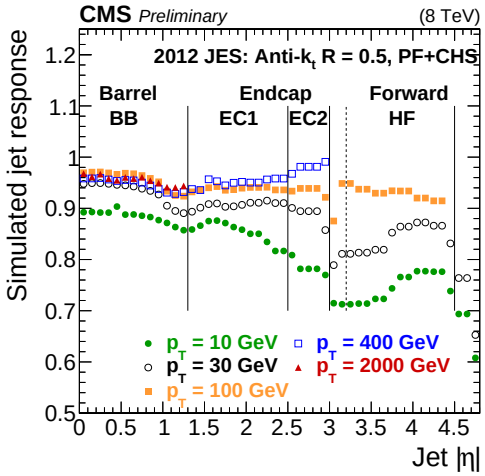


Figure 5.4: Jet response in simulation as a function of jet $|\eta|$ for different values of jet p_{T}. The inverse of the average response depending on jet η and p_{T} is used as correction factor [109].

The corrections are derived as follows: In a sample of simulated dijet events, the reconstructed jets are spatially matched to particle-level jets within $\Delta R < 0.25$. The simulated response is then determined as

$$R_{\mathrm{sim}} = \frac{p_{\mathrm{T,reco}}}{p_{\mathrm{T,ptcl}}}\,, \tag{5.2}$$

where $p_{\mathrm{T,reco}}$ and $p_{\mathrm{T,ptcl}}$ corresponds to the p_{T} of reconstructed and particle-level jet, respectively. The simulated response as a function of jet pseudorapidity for different jet p_{T} values is shown in Figure 5.4. The correction is subsequently determined as the inverse of the average measured response for each bin of jet η and p_{T}. The resulting correction is close to one for central jets with high p_{T}. For jets in the forward region and with lower transverse momenta, the correction factor can be as high as 1.65. This is due to effects such as radiative energy losses, which are especially strong for jets with high rapidities, and reconstruction imperfections due to the complex detector geometry in the endcap and forward region.

5.1.3 Data-Driven Corrections

As the simulation relies on model assumptions and approximations, additional data-driven jet corrections are mandatory. As these do not depend on any simulation program or physics model, the data-driven corrections are reliable and robust.

The data-driven calibration relies on balancing methods: In event topologies where a well-measured reference object is transversely balanced to a probe object, the momenta of both objects are related because of momentum conservation. The transverse momentum of the initial state is zero on average. From the precise measurement of the reference object, the momentum measurement of the probe object can be validated or corrected.

The CMS Collaboration performs the data-driven correction in two steps: a correction for the relative scale (differences between various detector regions in η) and the absolute scale.

The correction for the **relative scale** is derived from dijet balancing. The reference object is a jet in the central detector region where jet reconstruction is most precise because of the tracker, the high calorimeter resolution and the absence of complex transitions in the detector structure.

The tag object is a jet reconstructed in the endcap or forward region. In a balanced topology, where both jets are opposite of each other in ϕ, the absolute transverse momentum should be equal. From a possible imbalance between the two jets, i.e. a difference between the measured jet momenta, a correction factor can be extracted.

The data-driven calibration of the **absolute scale** is the last step, where a p_{T}-dependent correction for all jets is determined. This is achieved by combining the results from several calibration analyses: Z($\to \mu^+\mu^-$)+jet, Z($\to \mathrm{e}^+\mathrm{e}^-$)+jet, γ+jet and multijet studies.

The cross section for Z+jet events is much lower compared to dijet events. However, the centre-of-mass energy and luminosity of the LHC provide Z+jet events at a rate sufficient for adequate statistical precision. Given the small systematic uncertainties, Z+jet events allows to measure the jet response with high precision, perform independent cross-checks of the previous corrections steps and study systematic effects and biases.

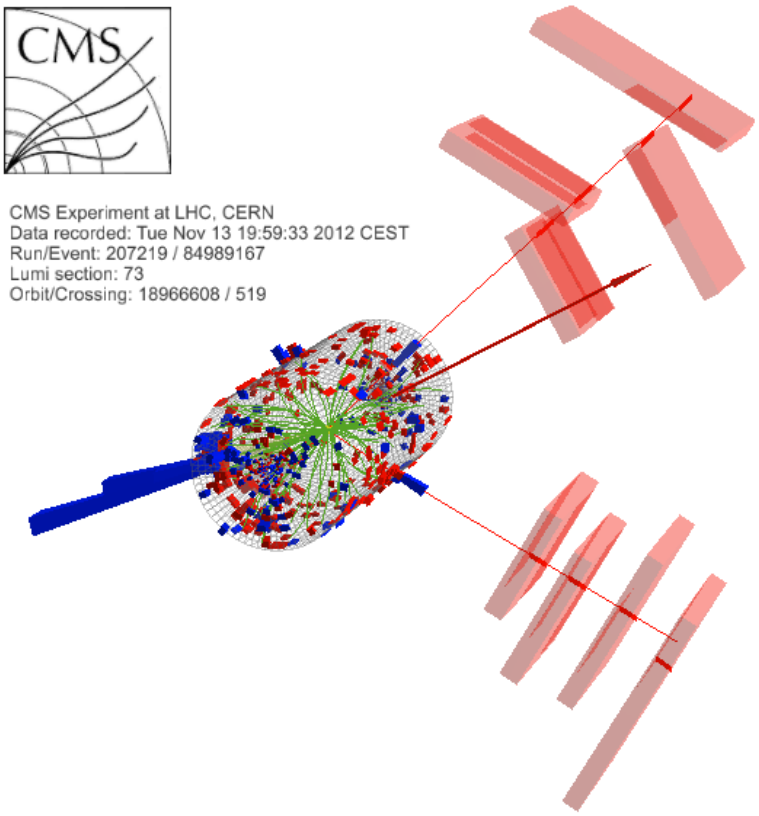
The Z($\to \mu^+\mu^-$)+jet channel assumes a special role as the reconstruction of the muons solely relies on information from the tracker and the muon chambers while the jet energy is measured in the calorimeters. Comparing the momenta of Z boson and jet, the detector components are effectively checked against each other. The Z($\to \mu^+\mu^-$)+jet calibration has been performed at KIT since 2008. The methods and results are detailed in [43, 59, 85, 96, 117–122].

The following sections detail the Z($\to \mu^+\mu^-$)+jet calibration studies. The combination of results from the different analyses and the determination of the data-driven corrections are explained at the end of this chapter.

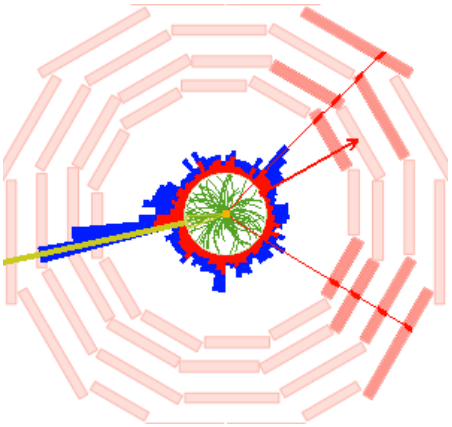
5.2 Event Selection

From the large number of collision events recorded by the CMS detector at 8 TeV, only events where a Z boson is balanced with a jet are suitable for calibration. An example of such a Z($\to \mu^+\mu^-$)+jet event is shown in Figure 5.5. Several selection steps, described in the following sections, are necessary to obtain a sample with only such events.

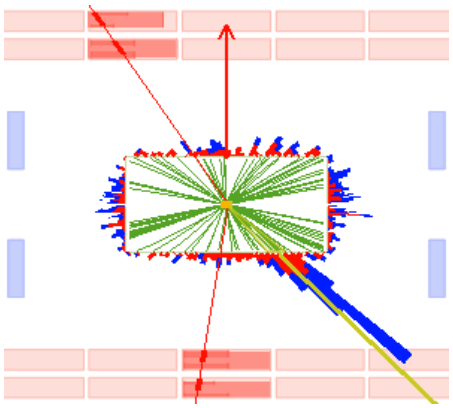
The processing of the data and the selection of events has been performed with a setup as described in Section 4.1.2, using Kappa, Artus and Excalibur. The graphs visualizing the results were produced with HarryPlotter and Merlin.



(a) 3D view



(b) Projection to the r-ϕ-plane



(c) Projection to the r-z-plane

Figure 5.5: Visualization of a Z($\rightarrow \mu^+\mu^-$)+jet event reconstructed with the CMS detector. The red tracks which extend to the outer detector layers, e.g. the muon chambers, belong to the two muons from the Z boson decay. The green lines are the particle tracks reconstructed in the tracker. The energy deposits in the calorimeters are shown in red (ECAL) and blue (HCAL). The yellow line indicates the momentum of the reconstructed jet. The central red arrow is the $E_{\mathrm{T}}^{\mathrm{miss}}$.
From the projection in the $r-\phi$-plane (lower left), one can see the balanced topology with the jet being opposite in ϕ to the muons respectively the Z boson. The numerous tracks and smaller calorimeter deposits are caused by pileup.

5.2.1 Trigger, Data Certification and Data Samples

The CMS Collaboration uses a chain of hardware and software triggers as described in Section 3.2.1. A trigger is used which selects events with at least one muon with p_T above 17 GeV and another one with p_T above 8 GeV. As the efficiencies for muon triggers are high and calibration analyses do not necessitate a detailed study on the absolute event yield, the effects of trigger efficiency differences between data and simulation can be neglected.

Events from time periods in which the CMS detector was not able to operate properly are excluded. This is done via a list of certified run periods centrally provided by the data certification group of CMS Collaboration. The recorded and certified data in 2012 correspond to an integrated luminosity of $19.712\,\text{fb}^{-1} \pm 2.6\%$ [73].

The data are compared to a sample of 10 million simulated Z/$\gamma^*(\rightarrow \mu^-\mu^+)$+Jet(s) events, normalized to the corresponding cross section of 1177.3 pb [18] computed with FEWZ [94] at next-to-next-to-leading-order. The matrix element of the hard process was calculated with MadGraph [93] while Pythia 6 [123] tune Z2* was used for the subsequent parton shower. The events were then passed to the detector simulation performed with Geant4 [95]. As a result, the simulation can be compared to data at the level of reconstructed physical objects.

The full list of technical details concerning data certification, CMS-internal names of data samples etc. is given in Appendix A.1.

5.2.2 Pileup Reweighting

The simulated sample should match the data in terms of pileup activity. However, to ensures adequate statistical precision in the tails of the pileup distribution, the simulated sample is generated with a higher fraction of events with a low and high number of vertices. Subsequently, the events in the simulated sample are weighted to match the measured pileup distribution in data.

From the measured instantaneous luminosity and the total soft proton-proton cross section, the expectation value for the number of pileup interactions μ can be calculated. Details on the luminosity measurement are given in Section 3.2.3. In simulation, μ is directly available as part of the information from the event generation. The events in the simulated sample are then weighted to match the distribution of expected pileup interactions in data. The distribution of the number of reconstructed primary vertices as a result of the weighting procedure can be seen in Figure 5.6.

Figure 5.7(a) shows the vertex reconstruction efficiency. The number of reconstructed vertices is shown as a function of the number of simulated vertices. A straight line is fitted to the data points. From the slope of the line, the pileup vertex reconstruction efficiency in Z($\rightarrow \mu^+\mu^-$)+jet events can be estimated as 0.65. The decrease in vertex reconstruction efficiency for a higher number of vertices is negligible – from the design specifications of the tracker, a significant decrease is not be expected until around 100 vertices. Because of this linear dependence, the number of reconstructed vertices is proportionate to the

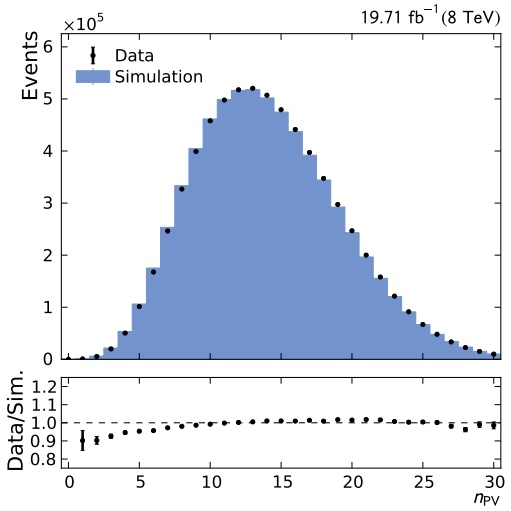


Figure 5.6: Distribution of number of reconstructed vertices. The expected pileup in data is calculated for each event from the inclusive proton-proton cross section and the instantaneous luminosity of the LHC. The simulation is weighted to match the distribution in data.

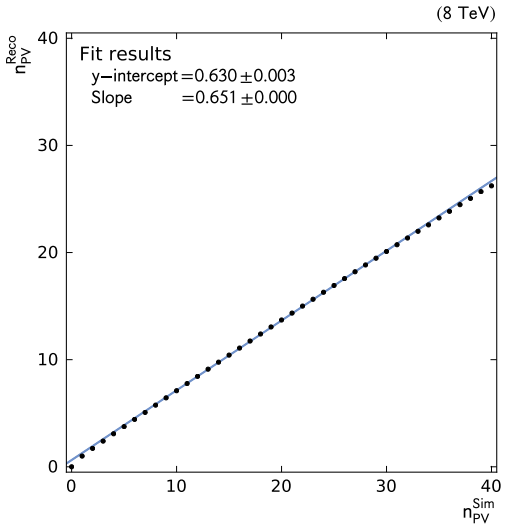
number of pileup vertices and can therefore be used as a rough estimate for the pileup activity in an event.

In Figure 5.7(b), the energy density ρ is shown as a function of the number of reconstructed vertices. The values for ρ scale linearly with N_{PV}. In the region of medium pileup, where the bulk of the events are located, data and simulation agree. Towards a very high or low number of vertices, the difference between data in simulation increases only slightly. This disagreement is not necessarily due to differences in the measured energy but could arise from small differences in the vertex reconstruction efficiency between data and simulation.

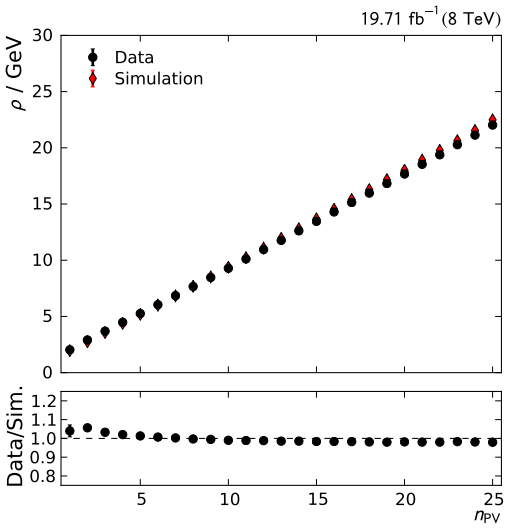
The expected and observed linear dependence between ρ and N_{PV} as well as the agreement between data and simulation show the success of pileup estimation with ρ and the pileup reweighting.

5.2.3 Jet and $\boldsymbol{E}_{\mathbf{T}}^{\mathbf{miss}}$ Corrections

The reconstruction and identification of jets with the CMS detector is described in Section 3.2.2, the correction procedure in Section 5.1. The studies in this chapter refer exclusively to jets reconstructed with the anti-k_T algorithm using a jet size parameter of $R = 0.5$, clustered from Particle-Flow candidates with Charged Hadron Subtraction (CHS, see Section 3.2.2) applied. As the $\mathrm{Z}(\rightarrow \mu^+\mu^-)$+jet analysis determines the absolute JES, the previous correction steps, e.g. the corrections for pileup, from simulation and the data-driven correction for the relative scale have to be already applied. As a basic reliability check, the actual correction factors for the different levels have to be studied.



(a) Number of reconstructed vertices as a function of the number of simulated vertices



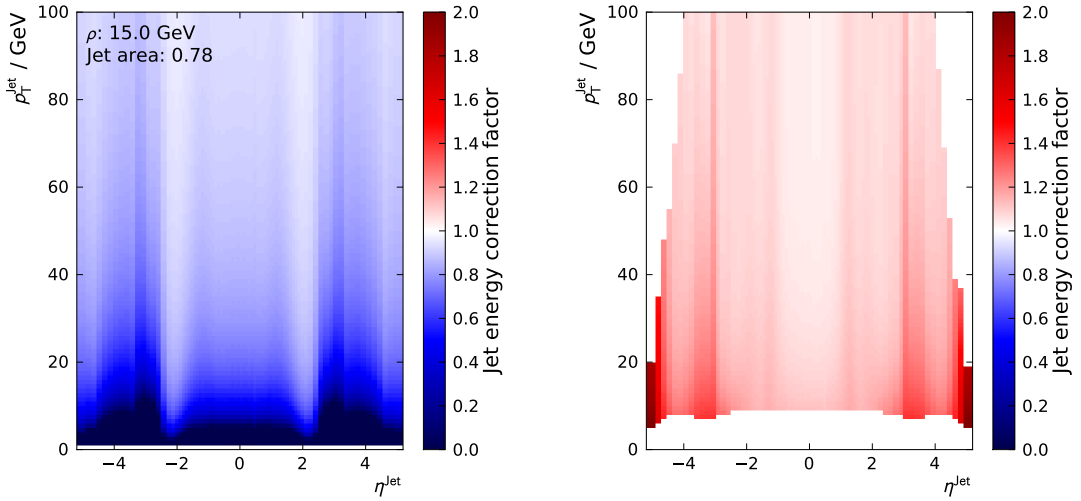
(b) Energy density ρ as a function of the number of reconstructed primary vertices

Figure 5.7: Left: Pileup vertex reconstruction efficiency. The number of reconstructed vertices in simulation is significantly lower than the number of true vertices. Fitting a straight line, the reconstruction efficiency can be estimated as around 0.65.
Right: The event energy density ρ is shown as a function of the number of reconstructed primary vertices in data and simulation.

The pileup correction factors are shown in Figure 5.8(a) depending on jet p_T and η for fixed values of jet area and ρ. $A_j = 0.78$ and $\rho = 15.0$ have been chosen as these values provide a good average for the selected events. A higher jet area or higher ρ would increase the pileup influence on the jet and decrease the correction factors. As this correction step subtracts the energy offset from pileup, the correction factors are below unity. The correction factors are especially small at low p_T, where the relative influence from pileup is largest, and in the detector region outside the tracker, i.e. $|\eta| > 2.4$, where pileup mitigation becomes more difficult.

The simulation corrections depend only on jet p_T and η. The correction factors, shown in Figure 5.8(b), increase at higher absolute pseudorapidity and lower transverse momentum to account for e.g. the increasing energy losses due to radiation. The differences for the correction factors between pseudorapidity regions are a result of the detector geometry: inhomogeneities or gaps between detector parts decrease reconstruction performance and thereby increase the necessary correction.

A similar effect can be seen for the data-driven corrections for the relative scale, see Figure 5.9: The correction factors also increase for higher absolute pseudorapidities, i.e. where the detector structure is more complex.However, only small corrections are needed for most regions, confirming the high quality and reliability of the detector simulation.



(a) Pileup correction

(b) Simulation correction

Figure 5.8: Applied jet energy correction factors for pileup (left) and simulation (right) corrections. As pileup corrections subtract the offset from pileup, the correction factors are below one and especially small in the low p_T and endcap region, where pileup influence on the jet is relatively high. The corrections from simulation are above one and especially high in the forward region, where jet reconstruction is particularly difficult.

E_T^{miss} Corrections

The purpose and determination of the missing transverse energy (E_T^{miss}) has been explained in Section 3.2.2. It is important not only as an experimental handle on weakly interacting particles, but also for studies on detector miscalibration, as will be explained in more detail in Section 5.4.

As E_T^{miss} is calculated from reconstructed objects, any changes of the momenta of these objects have to be propagated to E_T^{miss}, i.e. E_T^{miss} has to be recalculated. Such changes are introduced by the simulation corrections to jets which correct for effects such as non-linear detector response. The correction formula is:

$$\vec{E}_{T,corr} = \vec{E}_{T,raw} + \sum_{i}^{N_{jets}} \vec{p}^{\,i}_{T,raw} - \sum_{i}^{N_{jets}} \vec{p}^{\,i}_{T,corr} - \sum_{i}^{N_{jets}} \vec{\mathcal{O}}^{i}_{RC} \tag{5.3}$$

where $\vec{E}_{T,corr}$ and $\vec{E}_{T,raw}$ are the corrected and uncorrected E_T^{miss}, respectively. The transverse momenta of the jets are $\vec{p}_{T,raw}$ and $\vec{p}_{T,corr}$. $\vec{\mathcal{O}}^{i}_{RC}$ is the pileup offset calculated with the Random Cone method (see Section 5.1.1). With this approach, it is ensured that only the differences from the simulation corrections are taken into account and the pileup offset remains isotropic.

E_T^{miss} and its ϕ-direction after the correction can be seen in Figure 5.10. The E_T^{miss} is on average larger in data, emphasizing the need for an additional correction on the absolute JES.

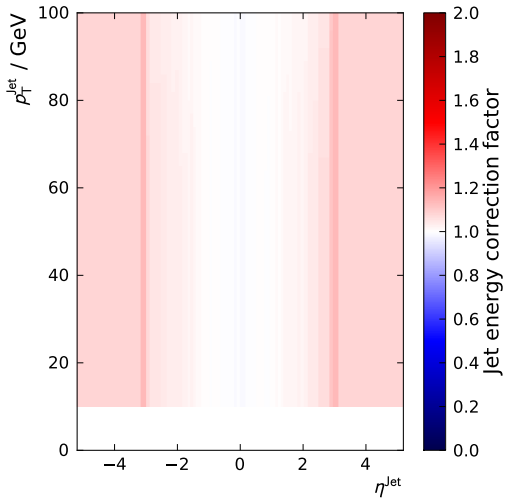


Figure 5.9: Jet energy correction factors for the relative scale in data, to be applied on top of pileup and simulation corrections. These corrections are small compared to the corrections from simulation (Figure 5.8), proving the reliability of the simulation. Jets with p_T below 10 GeV are not corrected, as the bias from pileup is too large for a reliable calibration.

The ϕ-distribution of the E_T^{miss} shows a sine-wave-like modulation which also is different for data and simulation. This is caused by a misalignment of the detector parts, particularly the calorimeters, in relation to the beam axis. As physics analyses average over ϕ, this effect does not cause any bias.

5.2.4 Z($\to \mu^-\mu^+$)+Jet Topology Selection

Further kinematic and topological selection steps have to be performed to obtain a sample which contains only Z($\to \mu^+\mu^-$)+jet events suited for calibration studies. The following selection criteria are applied:

Muon kinematic range The reconstruction and identification of muons with the CMS detector is detailed in Section 3.2.2. To ensures high precision, phase space regions where muon reconstruction is difficult (low energetic muons or muons reconstructed at the border of the tracker and the muon chambers) are excluded by requiring

$$|\eta_\mu| < 2.3\,, \tag{5.4}$$

$$p_{T,\mu} > 20\,\text{GeV}\,. \tag{5.5}$$

Z boson reconstruction The Z boson has to be reconstructed from a muon-antimuon-pair. Therefore, only events with at least two and a maximum of three muons (which have passed the previous kinematic criteria) are selected. In case of three muons, the muon-antimuon pair with the invariant mass closest to the Z boson mass (91.19 GeV [6]) is selected.

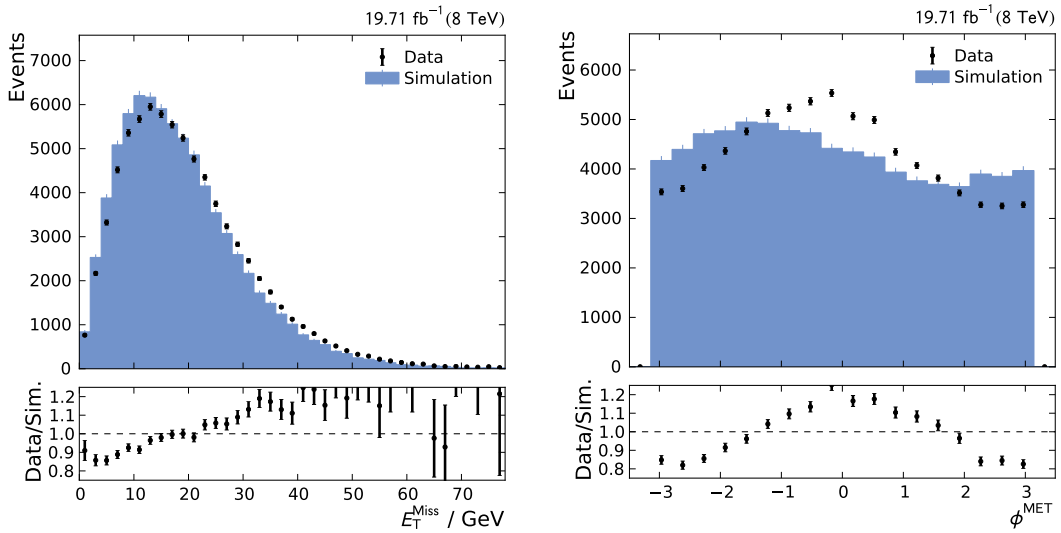


Figure 5.10: Missing transverse energy (left) and its direction in ϕ (right). The missing transverse energy is slightly higher in data. The ϕ-distribution, usually expected to be evenly distributed, shows a sine-wave-like modulation that is also different in data and simulation. It is caused by minor misalignment of the calorimeters with respect to the beam axis.

Dimuon invariant mass range In order to increase the probability of the selected muon pair to actually originate from a Z boson decay, only events are selected where the invariant mass of the dimuon system differs by less than 20 GeV from the nominal Z boson mass:

$$|m_{\mu\mu} - 91.19\,\text{GeV}| < 20\,\text{GeV}\,. \tag{5.6}$$

Z boson minimum p_T Jet calibration in low energetic phase space regions is not reliable because of large systematic biases, e.g. from pileup. Therefore, a minimum p_T threshold is introduced for the Z boson:

$$p_T^Z > 30\,\text{GeV}\,. \tag{5.7}$$

Jet minimum p_T Jets with low transverse momenta suffer from large out-of-cone or smearing effects. As these jets can not be reliably calibrated, a limit on the transverse momentum of the jet with the highest p_T – the *leading jet* – in each event is imposed:

$$p_T^{Jet} > 12\,\text{GeV}\,. \tag{5.8}$$

Jet maximum pseudorapidity In the endcap and forward region, jet reconstruction is more difficult because of the lack of information from the tracker, the complex detector geometry and the worse calorimeter energy resolution. Therefore, only events are considered where the leading jet is reconstructed in the barrel region:

$$|\eta_{Jet}| < 1.3\,. \tag{5.9}$$

Restriction on subleading jet momenta For the Z+jet calibration, a topology where a Z boson is balanced by exactly one jet is desired. In many cases, final state radiation leads to an additional jet, significantly decreasing the measured energy of the leading jet with respect to the parton energy. However, selecting events with only one jet is no viable solution: the pileup interactions in every event lead to a large background of low-energetic jets which cannot be reliably distinguished from radiation jets.

As a compromise, an upper limit on the p_T of the second-leading jet with respect to the Z boson p_T is introduced, limiting the possible energy loss for the leading jet due to radiation:

$$\frac{p_\mathrm{T}^{\mathrm{Jet2}}}{p_\mathrm{T}^{\mathrm{Z}}} < 0.2\,. \tag{5.10}$$

Azimuthal separation of jet and Z boson To further ensure a balanced topology, jet and Z boson are required to be opposite of each other in ϕ:

$$|\Delta\phi(\mathrm{Jet}, \mathrm{Z})| > 2.8\,. \tag{5.11}$$

The efficiency of the different selection steps ("cuts") is visualized in Figure 5.11. The aggregated efficiency of all selection steps is less than one percent, especially because of the restriction on subleading jet momenta.

The efficiency is different in data and simulation. The *Trigger* and *DataCert* filters select events according to the trigger decision and the official data certification, respectively. They are not applied in simulation and therefore have full efficiency.

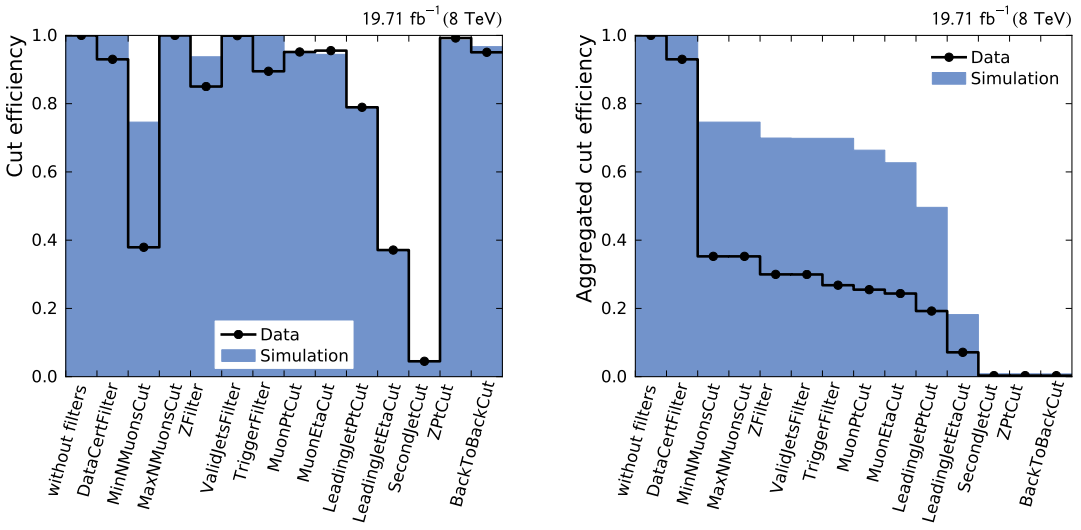
The difference for the *MinNMuonsCut* is due to the simulated sample containing only Drell-Yan events, while in data dimuon-signatures can originate from a number of physics processes. The rough preselection in data often contains reconstructed muons which do not fulfil the identification and kinematic selection criteria.

The second largest difference is in the *ZFilter*, which selects only events where a muon-antimuon pair with an invariant mass close to the nominal Z boson mass was found. In data, the muon-antimuon pair could originate from physics processes other than Z boson decay, e.g. $t\bar{t}$ production. Such events do not pass the ZFilter, thus the lower efficiency in data.

The second-leading jet maximum p_T requirement (*SecondJetCut)* has the overall lowest efficiency. This is due to more than one jet with considerable p_T in most events, either originating from final-state radiation or from pileup. Events with no radiation – which are perfectly suited for calibration studies – might not be selected because of such a pileup jet, showing the detrimental influence of pileup on the statistical precision.

5.3 Kinematic Distributions

For the calibration of the JES with balancing methods, the reference object, i.e. the Z boson reconstructed from the muons, has to be measured with high precision. As the



(a) Efficiency for each selection step.

(b) Aggregated selection efficiency

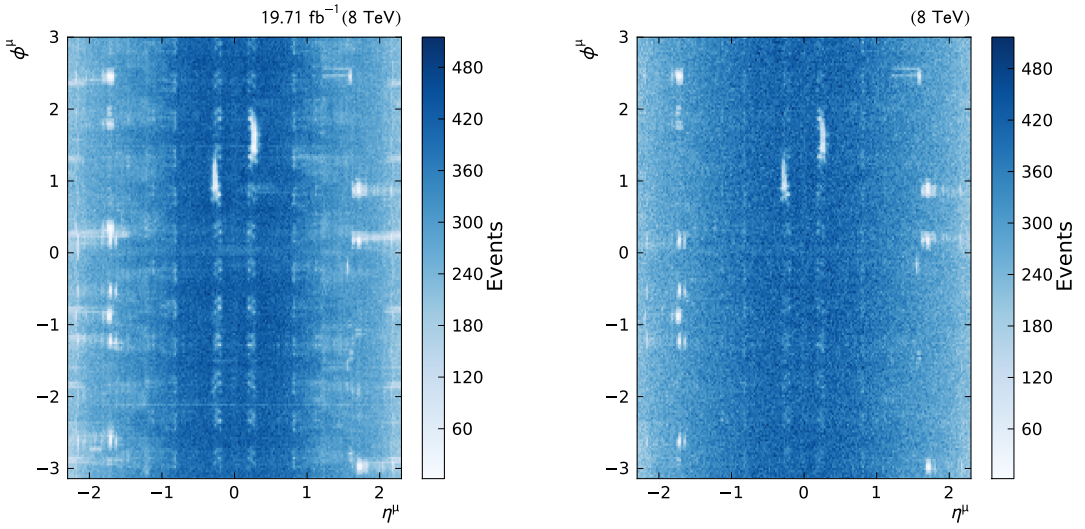
Figure 5.11: Efficiencies for the applied event selection criteria. After all selection steps, only a very small fraction (<1%) of the initial sample remains. The *SecondJetCut*, i.e. the maximum p_{T} requirement on the second-leading jet, has the lowest efficiency. The efficiency is different for data and simulation because of a different preselection.

correction on the absolute JES is derived from a comparison of data and simulation, the simulation has to reliably describe the data in all aspects necessary for a precise calibration, e.g. the properties of the involved physics objects (particles, jets, $E_{\mathrm{T}}^{\mathrm{miss}}$). To check whether these conditions are fulfilled, in this section the kinematic distributions are studied and the simulation is compared to data.

For calibration studies, the simulation has to describe the shape of a particular distribution, as one is interested in reconstruction performance. The predicted total event yield, which is important e.g. for cross section measurements, is not relevant. Effects like trigger or identification efficiency differences between data and simulation can be neglected. Therefore, all histograms for data-simulation comparisons in the following sections are normalized to the same number of events in data and simulation.

The distribution of muons in the η-ϕ-plane is shown in Figure 5.12. Several extraordinary features are visible: The horn-like features at $\eta \pm 0.2$ are the service chimneys for the magnet cryosystem. The patterns at high pseudorapidity are due to the difficult acceptance in that particular detector region, i.e. transitions between the chambers of the muon system. The irregular, asymmetric features throughout the η-ϕ-plane are caused by broken detector modules. However, these effects are known and included in the detector simulation. As a result, these features are reproduced in the simulation, see Figure 5.12(b).

Invariant mass and transverse momentum distributions of the Z boson reconstructed from the muons are shown in Figure 5.13. The shape of the kinematic distributions for



(a) Data (b) Simulation

Figure 5.12: η-ϕ distributions of muons in data (left) and simulation (right). The two "horns" in the upper middle belong to the cooling chimneys for the magnet cryostatic system. The regular patterns are due to detector geometry. The irregular spots at $|\eta| > 1.5$ are caused by faulty muon chamber modules. The distribution in data is well reproduced by the simulation, taking into account the detector effects and faults.

the Z boson are reproduced by the simulation. The position of the Z mass peak is nearly identical in data and simulation.

Figure 5.14(a) shows the rapidity of the Z boson. The distribution is symmetric for positive and negative values. The event count decreases towards higher values of the absolute rapidity, as is expected from the limited acceptance of the detector.

The Z bosons are evenly distributed in ϕ as is expected from symmetry principles, see Figure A.1 in Appendix A.2.1.

The reconstruction quality of the Z boson should not depend on the kinematic range. Therefore, it is important to check if the simulation describes the Z boson reconstruction also at different kinematic regions, e.g. at higher or lower transverse momentum. Figure 5.14(b) shows the average Z boson invariant mass as a function of the Z boson p_T. The mass increases at higher transverse momenta because of the larger available phase space. The differences between data and simulation are small, at the level of several per mille, confirming the high accuracy of the simulation.

The p_T and η distributions for the leading jet are shown in Figure 5.15. The jet p_T is lower in data, confirming the need for an additional data-driven correction of the JES. The effect of the $|\eta| < 1.3$ requirement is clearly visible.

The jets are reconstructed by applying a jet clustering algorithm to the collection of particles reconstructed with the Particle-Flow algorithm (see Section 3.2.2). The different

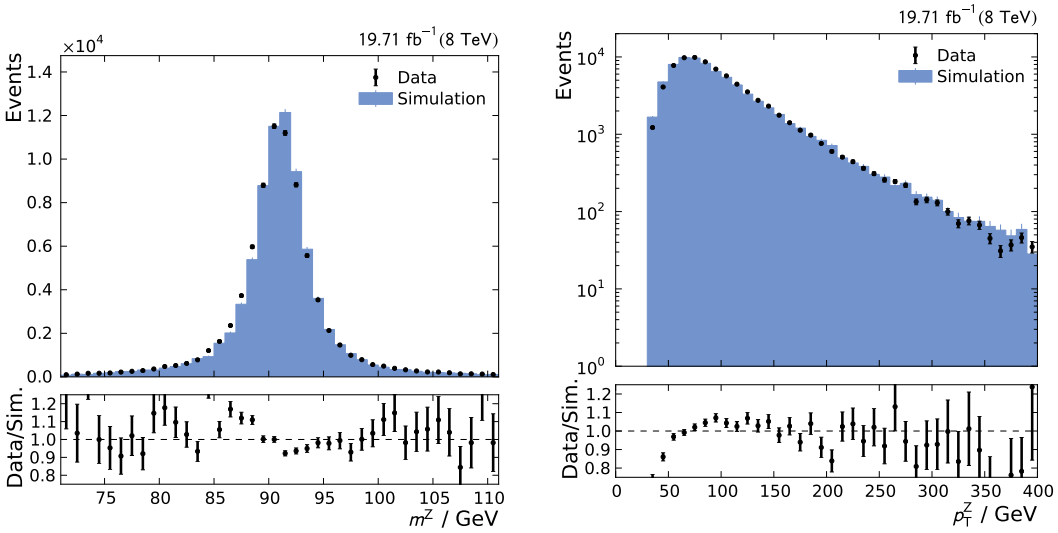


Figure 5.13: Invariant mass (left) and transverse momentum (right) of the dimuon system, i.e. the Z boson candidate. The measured distributions are well described by simulation.

categories of Particle-Flow candidates and their respective contributions to the measured jet energy are also studied to check the reliability of the Particle-Flow method.

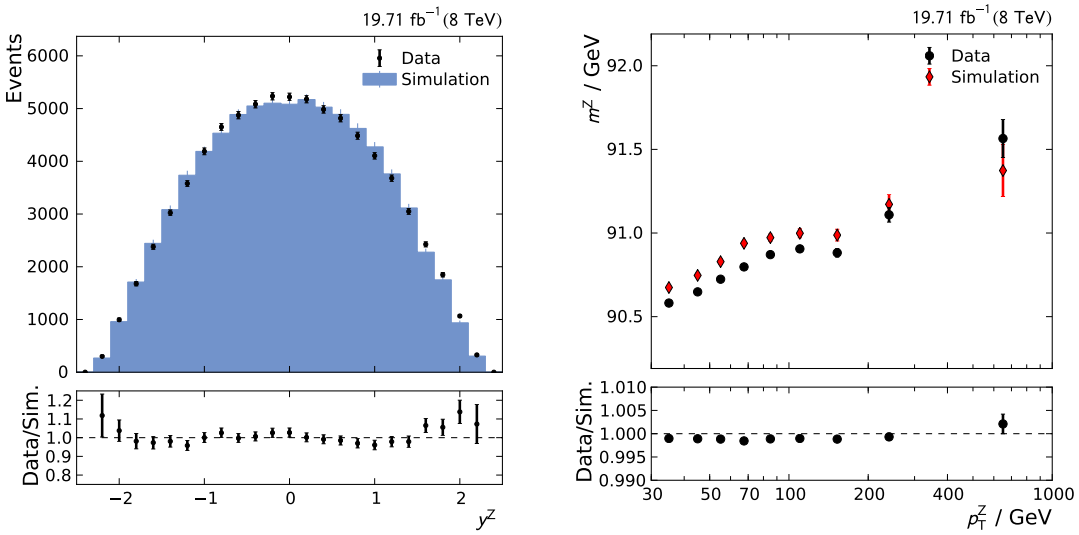
Differences in the Particle-Flow reconstruction between data and simulation could point to simulation imperfections or faults in detector components, as the different Particle-Flow categories correspond to specific detector parts (e.g. neutral hadrons are measured in the hadronic calorimeter).

The jet energy fractions for the Particle-Flow categories as a function of Z boson p_T are shown in Figure 5.16(a). The simulation agrees with the data for all components. The size of the different contributions is stable with respect to the p_T range, proving the success of the calorimeter calibration for different energy ranges.

Figure 5.16(b) shows the Particle-Flow contributions for different pseudorapidity ranges. The differences between the detector regions are strong as the detector structure and therefore the particle reconstruction varies. Within the tracker coverage, i.e. for $|\eta| < 2.4$, the simulation describes the contributions from the different categories well. In the endcap and forward region outside the tracker, where jet reconstruction is more difficult, the differences are slightly larger. In the forward region, the energy measurement is performed almost exclusively by the hadron forward calorimeter.

The contributions to the jet energy for the different Particle-Flow categories are shown in Figure 5.16(b). The contributions increase for a higher number of vertices as is expected from the additional pileup. The size of the measured contributions is well described by simulation.

Further comparisons are shown in Appendix A.2.1. As the basic kinematic quantities have been checked and are reliably described by simulation, calibration studies can be performed.



(a) y_Z (b) Z boson mass dependence on p_T

Figure 5.14: Left: Rapidity of the Z boson candidate. The simulation prediction agrees with the data. Right: Z boson invariant mass as a function of p_T.The average mass increases at higher p_T because of the larger available phase space. The differences between data and simulation are at the level of only few per mille.

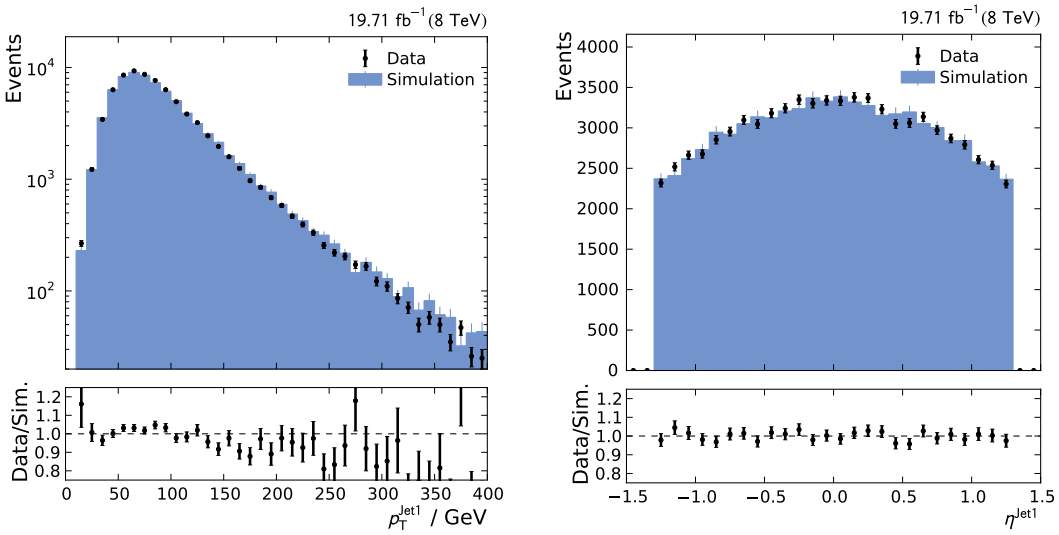
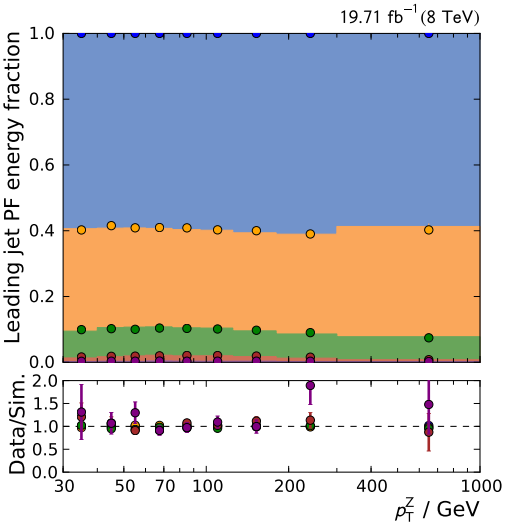
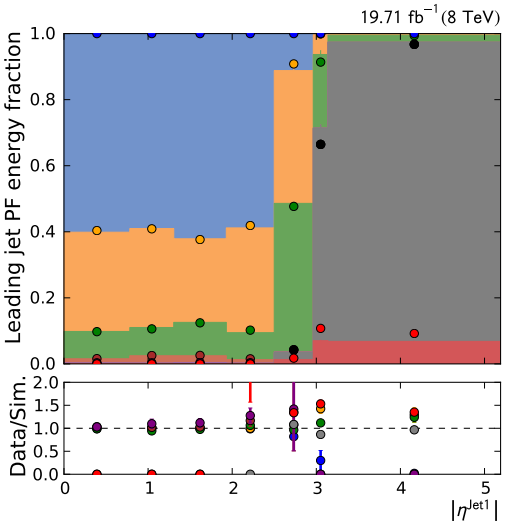


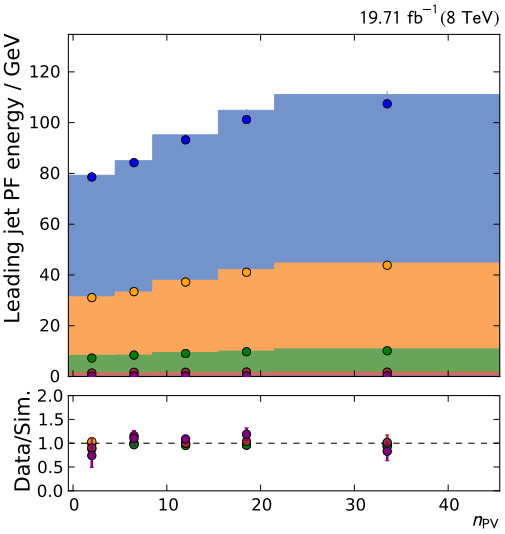
Figure 5.15: p_T (left) and η (right) distributions for the leading jet.



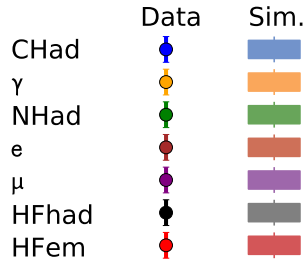
(a) PF composition as a function of p_T^Z.



(b) PF composition as a function of leading jet $|\eta|$.



(c) PF energy contributions as a function of the number of reconstructed vertices.



(d) Legend for PF categories. CHad and NHad refer to charged and neutral hadrons, respectively.

Figure 5.16: Particle-Flow (PF) composition of the leading jet. The PF composition is stable as a function of p_T^Z (upper left), confirming the reconstruction performance to be independent of the energy range. As a function of jet pseudorapidity (upper right), the PF composition changes because of the different detector structure. The energy contributions from the different PF categories as a function of the number of reconstructed primary vertices are shown in the lower left. They contributions increase with the number of vertices as is expected from the pileup influence.

5.4 Jet Response Measurements

The jet energy calibration relies on the comparison of the measured with the particle-level jet energy. This relation is expressed in terms of the *jet response*. In simulation, it can be simply calculated by comparing the p_T-ratio of a reconstructed and the matched ($\Delta R < 0.5$) particle-level jet, see Equation 5.2.

For the determination of the jet response in data, more elaborate methods are necessary. Two complementary techniques are used: the p_T balance and the MPF method. Both rely on the fact that the transverse momentum in the initial state is zero on average (minor transverse momentum arises only from the intrinsic parton momenta) and that the kinematic quantities of Z boson and jet are related via momentum conservation. Both methods are complementary and used for cross-checking their respective results.

p_T balance In balanced Z+jet events, one can determine the jet response simply via the p_T ratio between jet and Z boson:

$$R_\mathrm{balance} = \frac{p_\mathrm{T}^\mathrm{Jet}}{p_\mathrm{T}^\mathrm{Z}}\,. \tag{5.12}$$

This method produces a robust estimate of the jet response. However, final-state radiation can lead to a splitting of the leading jet and greatly reduce the measured jet momentum with respect to the original parton momentum, biasing the p_T balance towards lower values. This bias is weakened by the restriction on subleading jet momenta (see Equation 5.10), but cannot be completely avoided because of the large background from pileup jets.

Missing E_T Projection Fraction (MPF) An alternative, rather indirect approach is followed: At the particle level, the momentum vector sum of Z boson and jet is zero:

$$\vec{p}_\mathrm{T}^\mathrm{Z} + \vec{p}_\mathrm{T}^\mathrm{\,jet} = 0\,. \tag{5.13}$$

As there is no intrinsic missing energy in Z $\to \mu^+\mu^-$ decays, i.e. no neutrinos in the final state, this can be written as

$$R_Z\ \vec{p}_\mathrm{T}^\mathrm{Z} + R_\mathrm{jet}\ \vec{p}_\mathrm{T}^\mathrm{\,jet} = -\vec{E}_\mathrm{T}^\mathrm{miss}. \tag{5.14}$$

with the response of the Z boson R_Z and the jet response R_jet.

Combining equations 5.13 and 5.14 yields

$$R_\mathrm{jet} = R_\mathrm{Z} + \frac{\vec{E}_\mathrm{T}^\mathrm{miss}\cdot\vec{p}_\mathrm{T}^\mathrm{Z}}{\left(p_\mathrm{T}^\mathrm{Z}\right)^2}. \tag{5.15}$$

As the Z boson can be measured with high precision, the Z boson response can be assumed to be equal one and Equation 5.15 can be written as

$$R_\mathrm{MPF} := R_\mathrm{recoil} = 1 + \frac{\vec{E}_\mathrm{T}^\mathrm{miss}\cdot\vec{p}_\mathrm{T}^\mathrm{\,Z}}{\left(p_\mathrm{T}^\mathrm{Z}\right)^2}\,. \tag{5.16}$$

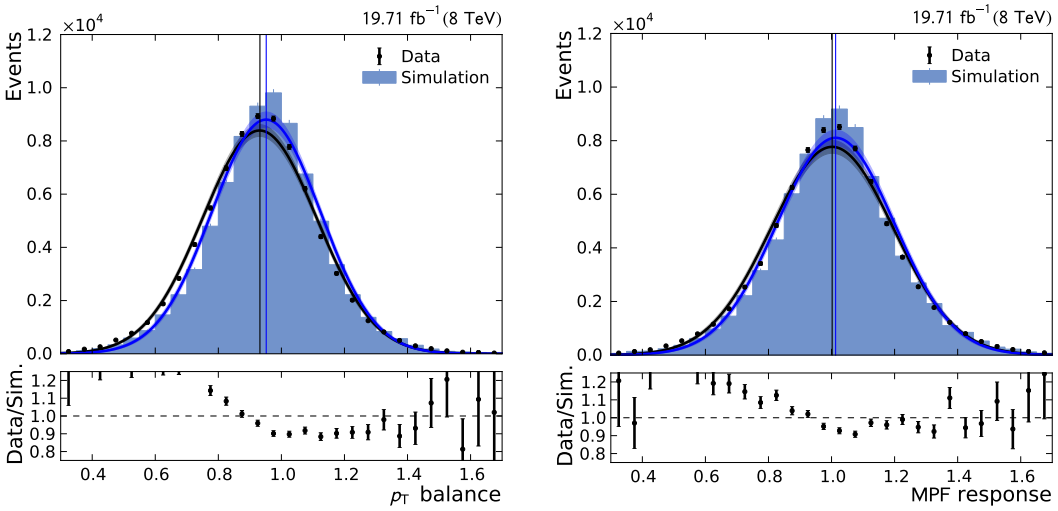


Figure 5.17: Distributions for p_T balance (left) and MPF response (right). A Gaussian has been fitted to the distributions in data and simulation. The vertical lines indicate the mean of the distribution. The width of the distribution is slightly larger in data for both methods, however the mean values agree between data and simulation. The average response is slightly lower with the p_T balance method because of the bias from final-state radiation.

where the result is defined as the MPF Response R_{MPF} , the jet response determined with the MPF method [108]. The MPF method relies on $E_{\text{T}}^{\text{miss}}$ and thereby on information from all detector subsystems. Therefore, it can only be used if the detector is already well calibrated.

In contrast to the p_T balance method, the MPF response is not sensitive to final-state radiation: If jet splitting occurs, the energy of the second jet still enters the calculation through $E_{\text{T}}^{\text{miss}}$. On the other hand, the MPF method is more biased by pileup, which can be non-isotropic in an event and affect the calculation of $E_{\text{T}}^{\text{miss}}$. However, because of the random pileup distribution, this bias cancels out on average.

The results from both methods are shown in Figure 5.17. Both distributions are Gaussian-shaped. The width of the distribution is slightly larger in data for both methods because of the worse resolution in particle and jet reconstruction. For the MPF method, the mean of the distribution is slightly above 1.0, while for p_T balance the mean is around 0.95. The lower average value for the p_T balance is due to the above-mentioned energy loss by final-state radiation.

A comparison of each method with the simulated response is shown in Figure 5.18. The correlation of p_T balance with the simulated response is higher than for MPF. This is due to pileup effects biasing the $E_{\text{T}}^{\text{miss}}$ distribution, an effect which however cancels on average. The p_T balance measurement suffers from final-state radiation, which biases the distribution systematically towards lower values.

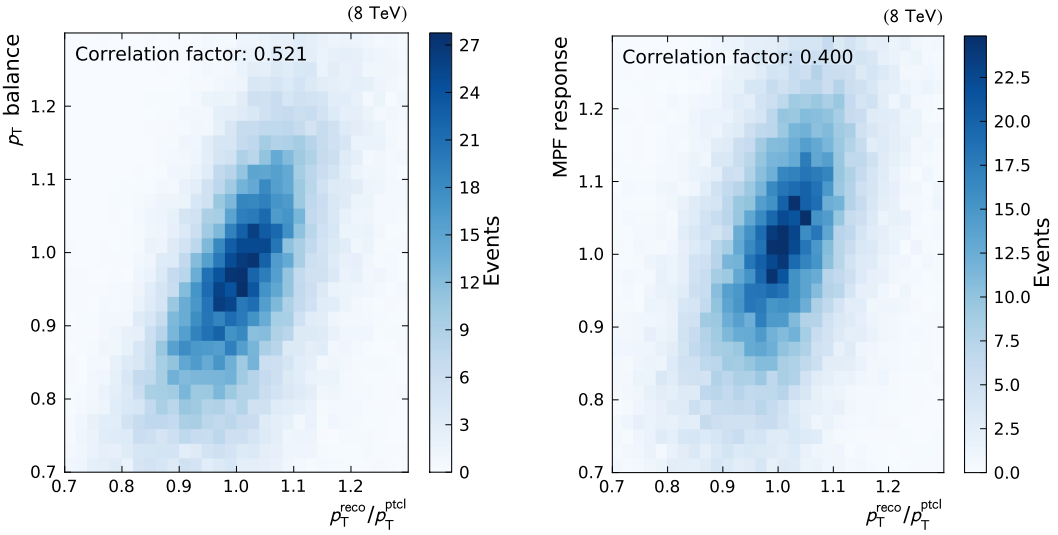


Figure 5.18: Two-dimensional distributions of the jet response from different methods. The correlation between simulated response and p_T balance (left) respectively MPF response (right) is shown.

Figure 5.19 shows the correlation between the p_T balance and MPF response. The correlation is rather low because of the pileup bias affecting MPF but not p_T balance in an event. Similarly, the bias from final-state radiation might affect p_T balance but not MPF.

These effects can be further examined by studying the root mean square (RMS) of the relative difference between the simulated response and p_T balance respectively MPF response. This value is calculated as

$$RMS = \sqrt{\frac{1}{N}\sum^{N}\left(\frac{R_{\text{sim}} - R}{R_{\text{sim}}}\right)^2} \tag{5.17}$$

with R being either the p_T balance or MPF response and the number of events N. It is determined in bins of p_T^Z and N_{PV} to obtain an estimate of how well each method approximates the simulated response for different regions of pileup influence and p_T, see Figure 5.20.

For the MPF response, there is a trend towards lower RMS values at higher p_T, likely due to the decreasing influence of pileup. The RMS values for the p_T balance are generally stable as a function of p_T^Z and N_{PV}. At $p_T \approx 150\,\text{GeV}$, the two methods exhibit similar performance. These results show the advantages of each method and further emphasize the need for a combined usage to achieve supreme precision.

To cross-check the results of the previous correction steps as well as to detect any discrepancies in particular phase space regions, the jet response has to be determined for different values of p_T, different detector regions of jet reconstruction and different pileup scenarios.

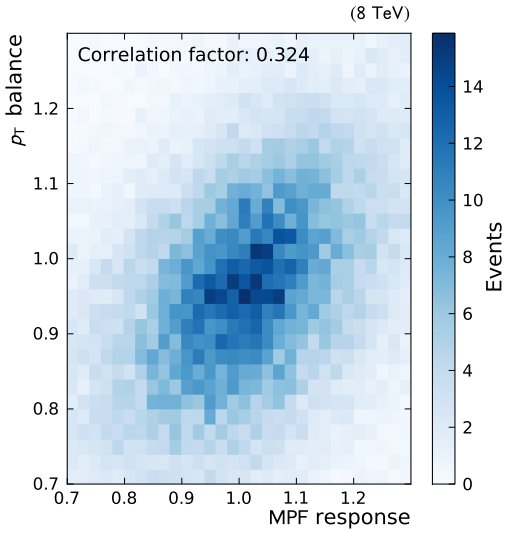


Figure 5.19: Correlation between p_T balance and MPF response. Even though both methods are used to measure the jet response and agree on average, the correlation for a single event between them is rather low.

p_T balance and MPF response as a function of Z boson p_T are shown in Figure 5.21. A systematic difference between data and simulation can be observed, again confirming the need for a further correction of the absolute JES. The results from the p_T balance method are lower compared to the MPF response and decrease further at lower p_T, likely due to final-state radiation.

Figure 5.22 shows the jet response as a function of the absolute pseudorapidity of the jet. For the p_T balance, a continuous decrease can be observed towards higher values of $|\eta|$. Again, this is likely caused by energy loss through final-state radiation which increases at higher rapidities. With the MPF method, a small increase can be seen in the region $1.3 < |\eta| < 3$.

To check the results of the pileup corrections, the pileup dependence of the jet response is studied, quantified by the number of reconstructed primary vertices N_{PV}. The result is shown in Figure 5.23. For both methods, only a minor dependence on N_{PV} can be observed. The jet response difference between the bins with lowest and highest pileup is around (1 to 2) %. Again, the values obtained with p_T balance are systematically lower compared to the MPF response and a minor difference between data and simulation can be observed.

The performance of the CMS detector can decrease over time, mainly because of calorimeter radiation damage, especially in the endcap and forward region where particle flux is highest. In previous studies [43, 119], a significant time dependence of the calorimeter response in the endcap section has been observed. With recalibration of the calorimeter signals in data, these effects need to be rechecked.

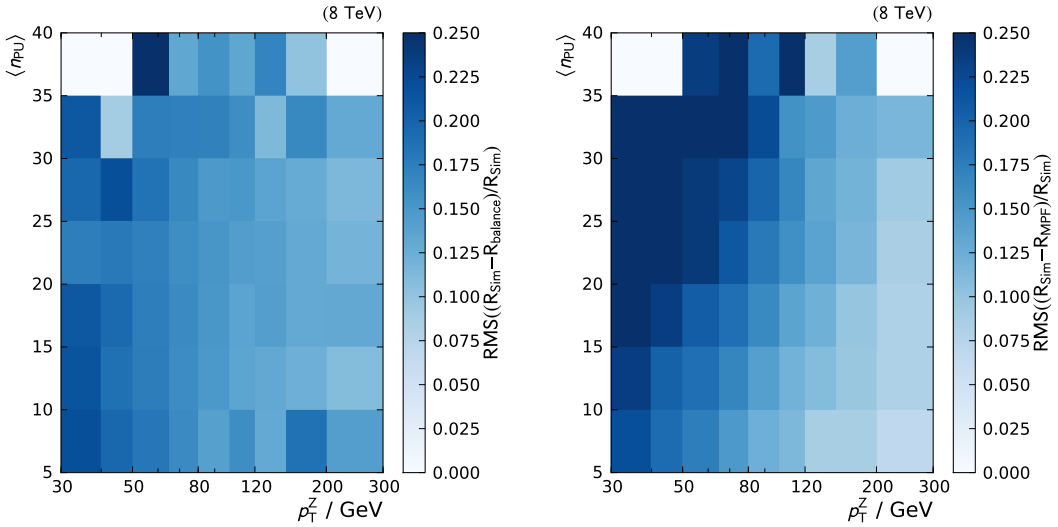


Figure 5.20: The RMS for the relative difference between the simulated response and the p_T balance (left) or the MPF response (right) is calculated as a function of p_T^Z and N_{PV}. The RMS for the MPF method shows a larger dependence on p_T and N_{PV} as the MPF method is more affected by pileup effects, especially at low p_T.

The jet response is studied as a function of time, see Figure 5.24. For the p_T balance, the jet response deteriorates by a few percent in the endcap section. However, this effect can not be reliably confirmed by the corresponding measurement for the MPF response.

The various cross-checks performed in this section have shown that the jet response can be reliably measured with both the p_T balance and the MPF method. Both have their advantages and are affected by different biases. For maximum precision, the results from both methods are combined. This is done via a global fit in the combination of all channels which will be described in Section 5.8.

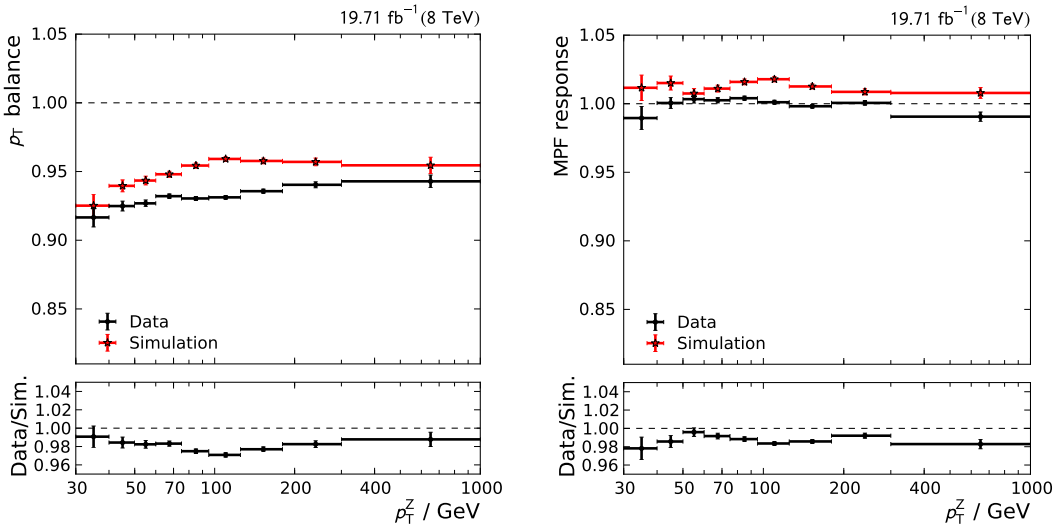


Figure 5.21: Jet response as a function of Z boson p_T for the p_T balance (left) and MPF method (right).

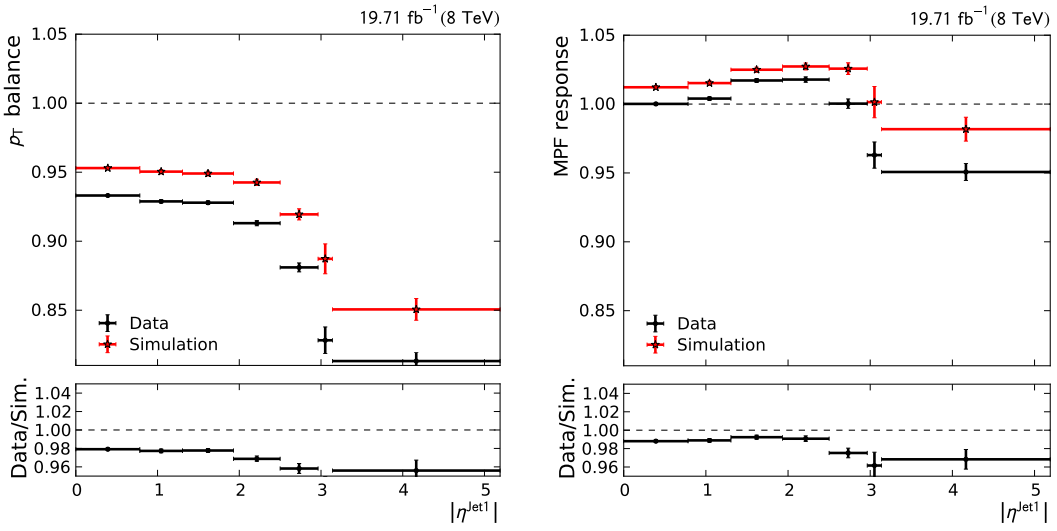


Figure 5.22: Jet response as a function of absolute pseudorapidity of the leading jet for the p_T balance (left) and MPF method (right). The p_T balance decreases toward higher values of $|\eta|$ because of higher radiative losses. Here, the $|\eta_{\text{jet}}| < 1.3$ requirement was not applied.

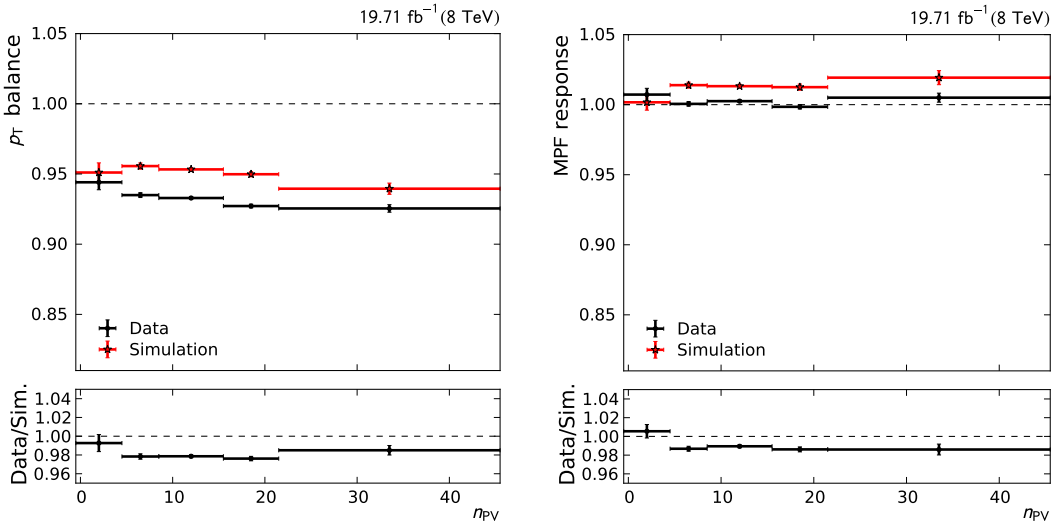


Figure 5.23: Jet response as a function of the number of reconstructed primary vertices for the p_T balance (left) and MPF method (right). The more vertices per event, the larger the effects from pileup. The jet response is stable as a function of N_{PV}, showing the success of the pileup corrections.

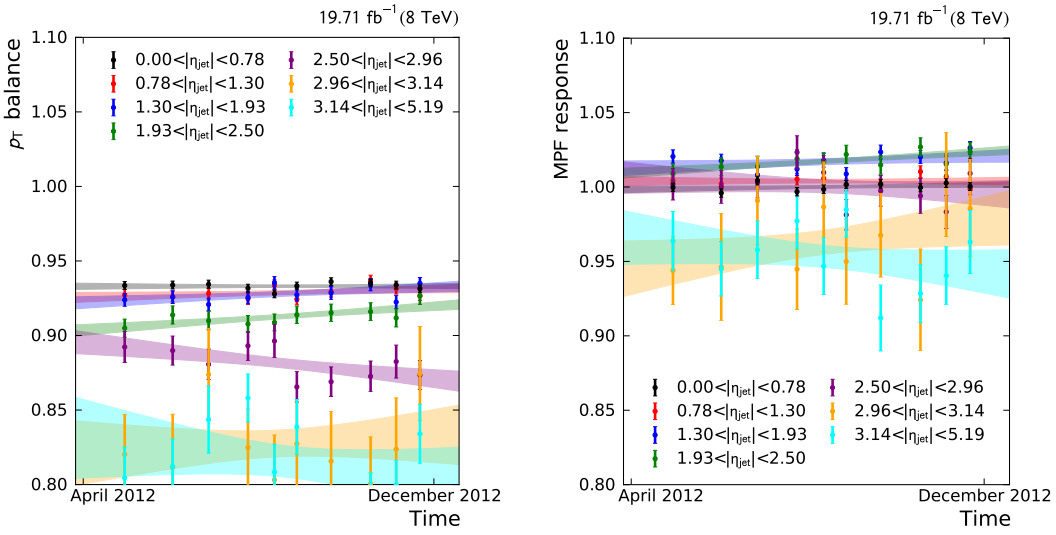


Figure 5.24: Time dependence of the jet response measured with p_T balance (left) and MPF (right). Small, but significant time dependence can be observed with the p_T balance method in the endcap region of the detector.

5.5 Jet Flavour Studies

Certain properties of a jet depend on the flavour of the originating parton: Reconstructed jets from bottom or charm quarks can lose energy via neutrinos from semileptonic b or c hadron decays. Gluon jets have a wider fragmentation shape, increasing out-of-cone effects [8]. The impact of these effects has to be carefully studied to reduce the resulting uncertainty on the JES.

In the following sections, jet response differences between flavours are determined and experimental results are compared with simulation. Multivariate tagging techniques are employed for the difficult task of jet flavour determination in data.

5.5.1 Jet Flavour in Simulation

Even with the in-depth information in simulation, the flavour of a jet cannot be unambiguously defined. The definition depends on which parton is regarded as determining the jet's properties, which in turn depends on the choice of the simulation program and the hadronization scheme.

Various procedures to determine the jet flavour in simulation are used by the CMS Collaboration [124]. In the studies presented here, the *physics definition* is used. It is based on the particle status codes of the Pythia 6 [123] Monte Carlo event generator. Reconstructed jets are matched to partons of the *initial state* (status=3 in Pythia 6). A match is defined as ΔR between jet and parton being smaller than 0.3. The flavour of the matching parton is assigned to the jet. If there is not exactly one matching parton (i.e. more than one or no match), the jet flavour remains undefined.

The flavour distribution of the leading jet in $\mathrm{Z}(\rightarrow \mu^+\mu^-)$+jet events is shown in Figure 5.25. The distribution corresponds to the parton distribution in the proton, see Figure 2.5. Down- and up-type jets occur more often than their antiquark counterparts, with the difference being roughly twice as big for u jets than for d jets. This reflects the valence quark composition (two u quarks and one d quark) of the proton. Heavy quark jets are less common. Gluon jets occur most often, corresponding to the large gluon PDF at high scales.

The flavour distribution changes as a function of Z boson p_T and jet pseudorapidity, see Figure 5.26. This dependence originates from the relation of the proton PDFs with the kinematic quantities of the Z+jet pair: Equation 3.5 shows that for a high absolute rapidity, the difference between x_1 and x_2 must be large, i.e. one of the involved partons has to carry a rather low fraction of the proton momentum. Thereby, jets in the forward region probe the proton at lower x, where the gluon PDF increases.

High transverse momenta of jet and Z boson require high x_1 and x_2. Therefore, at higher transverse momentum the flavour composition is determined by the PDFs at high x, which are dominated by the valence quark PDFs.

The fraction of jets with undefined flavour increases at lower p_T because of out-of-cone and smearing effects, which shift the four-vector of the reconstructed jet with respect to the particle-level jet and thereby impede the parton-jet matching. At higher momenta, additional partons from initial-state radiation might emerge in a direction close to the

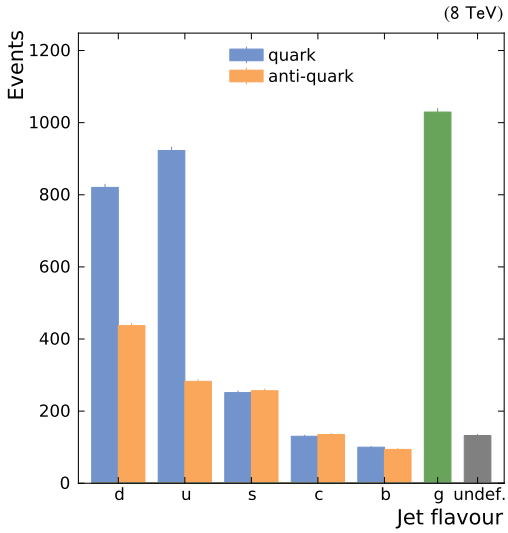


Figure 5.25: Flavour of the leading jet in simulation. The flavour distribution corresponds to the PDFs of the proton (see Figure 2.5) and the Z+jet production channels (see Figure 2.2).

leading jet. With more than one parton fulfilling the $\Delta R < 0.3$ condition, the jet flavour is undefined with the physics definition.

Especially for data-driven studies, where particle-level information is not available, it is important to determine how flavour effects influence the properties of the reconstructed jet. This can be estimated by studying the Particle-Flow composition depending on jet flavour, see Figure 5.27.

For heavy flavours (i.e. b and c quarks), the energy contribution from electrons and muons from semileptonic decays of c and b hadrons is visible. The higher fraction of neutral hadron energy for s quarks originates from the higher probability of s quarks to fragment into neutral hadrons such as kaons.

Figure 5.28(a) shows the jet response for different jet flavours. p_T balance, MPF response and simulated response are compared. MPF and p_T balance decrease for c and even more for b jets because of neutrinos from semileptonic decays, which lead to energy loss in the reconstructed jet. The ratio $p_\mathrm{T}^\mathrm{reco}/p_\mathrm{T}^\mathrm{ptcl}$ is not affected by this as the neutrino energy is missing in both $p_\mathrm{T}^\mathrm{reco}$ and $p_\mathrm{T}^\mathrm{ptcl}$. The response for s quark jets is lower compared to jets from u or d quarks. This is due to s quarks tending to fragment into neutral hadrons, which have a slightly lower reconstruction efficiency compared to other Particle-Flow categories. For gluon jets, p_T balance yields lower values because of the fragmentation behaviour of gluons: the fragmentation into a wider cone and many low-energetic particles leads to more out-of-cone energy loss and decreased reconstruction efficiency.

The p_T balance for different flavours and as a function of $p_\mathrm{T}^\mathrm{Jet2}/p_\mathrm{T}^\mathrm{Z}$ is shown in Figure 5.28(b). Events with neutrinos are excluded to suppress the decrease of the jet response because of neutrinos escaping the detector unnoticed. Extrapolating the results

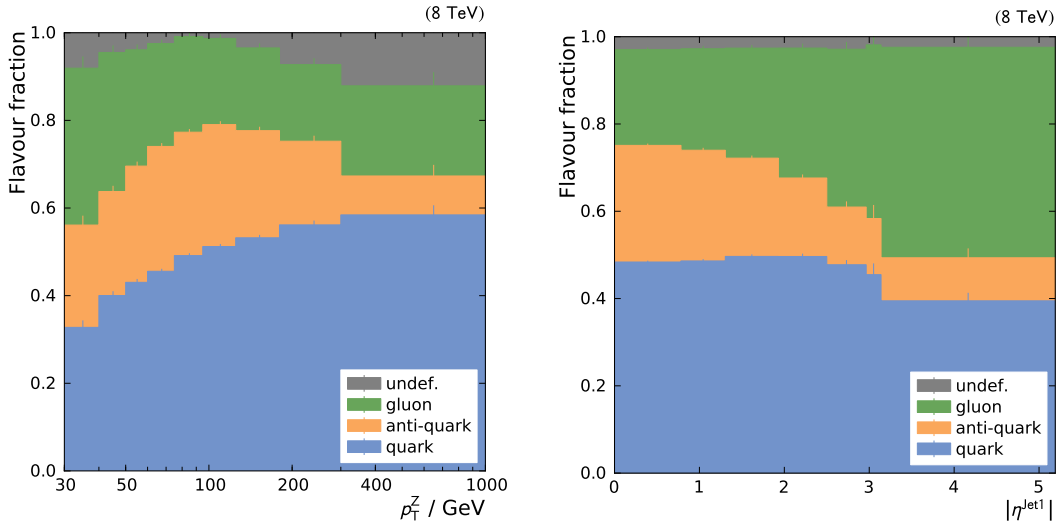


Figure 5.26: Flavour fractions as a function of p_T^Z (left) and absolute jet pseudorapidity (right). The fraction of gluon jets increases towards lower p_T and higher rapidity. The fraction of jets from quarks increases towards higher p_T, especially compared to the fraction of jets from antiquarks.
These effects are caused by the parton distribution inside the proton, see Figure 2.5. For high p_T, the high-x region of the proton is probed, corresponding to higher PDFs for the valence quarks. At higher rapidity, the flavour distribution corresponds to the PDFs at low x, e.g. a higher gluon PDF.
The fraction of jets with undefined flavour is highest at low p_T, where no parton within $\Delta R < 0.3$ with respect to the jet can be found, and at higher p_T, where there is more than one parton close to the jet because of additional partons from initial-state radiation.

to $p_T^{Jet2}/p_T^Z \to 0.0$, the response values for all flavours converge. This shows that (in the absence of neutrinos) the observed differences between flavours are mostly caused by different fragmentation behaviour and vanish at the perfect event topology with only one jet.

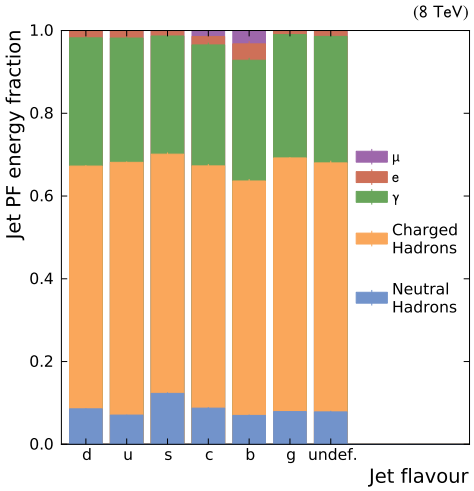
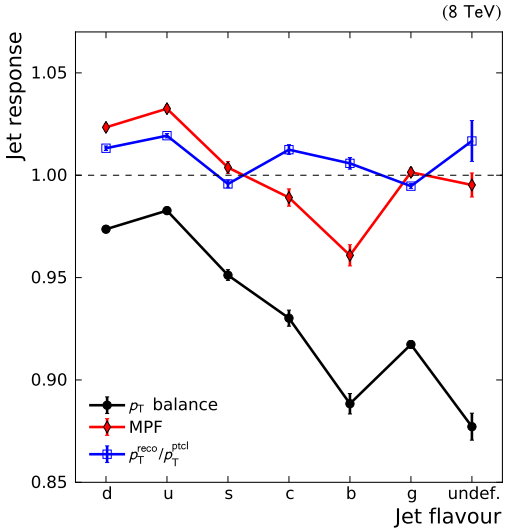
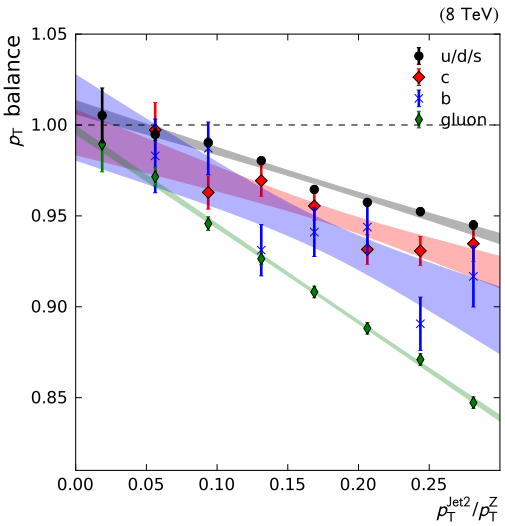


Figure 5.27: Particle-Flow jet composition for different jet flavours. The energy fraction from electrons and muons is higher for c and b jets because of semileptonic decays of c and b hadrons. The neutral hadron fraction is higher for s-jets because of s quark fragmentation into kaons.



(a) Jet response for different methods and different jet flavours



(b) p_T balance as a function of $p_\mathrm{T}^{\mathrm{Jet2}}/p_\mathrm{T}^{\mathrm{Z}}$ for different flavours

Figure 5.28: Left: p_T balance, MPF response and simulated response ($p_\mathrm{T}^{\mathrm{reco}}/p_\mathrm{T}^{\mathrm{ptcl}}$) for different jet flavours. Various systematic effects (decays into neutrinos, shower shape, fragmentation behaviour) result in differences in the jet response between flavours.
Right: p_T balance as a function of $p_\mathrm{T}^{\mathrm{Jet2}}/p_\mathrm{T}^{\mathrm{Z}}$ for different flavours, excluding events with neutrinos. The p_T balance converges for all flavours at $p_\mathrm{T}^{\mathrm{Jet2}}/p_\mathrm{T}^{\mathrm{Z}} = 0$ to 1, proving that the measured response differences are mostly caused by radiation effects and vanish at the perfect event topology of only one jet.

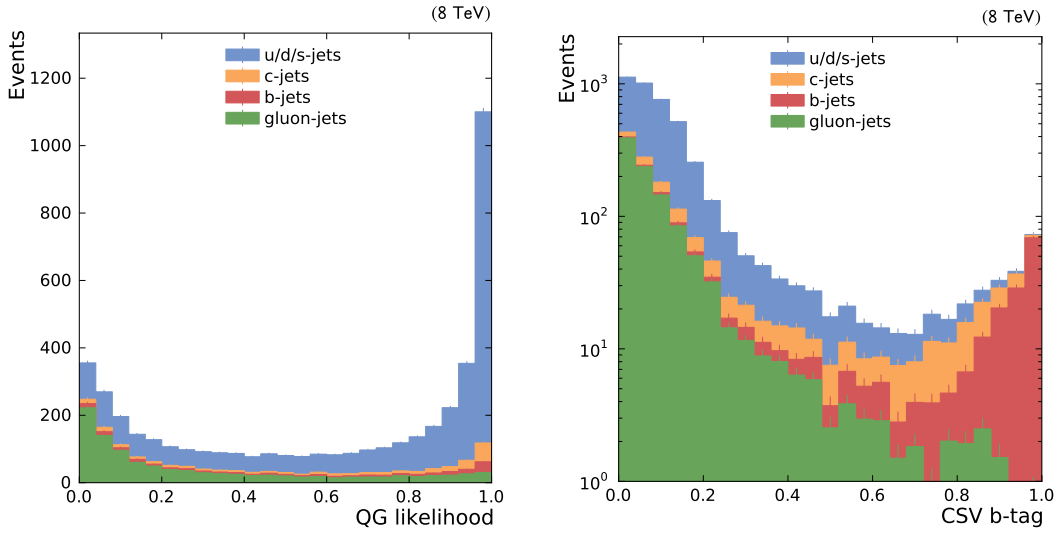


Figure 5.29: Distribution of gluon (left) and b tagging (right) discriminators in simulation. The jet flavour has been determined from simulation with the physics definition. Both tagging methods succeed in separating gluon respectively b jets from other flavours.

5.5.2 Measuring Flavour Dependence with the 2D-Tagging Method

The determination of the jet flavour via jet-parton matching is of course only possible in simulation. In data, only information on the reconstructed jet is known, complicating jet flavour studies.

To cope with this challenge, sophisticated *tagging* methods have been developed. These methods calculate the probability of the jet to originate from a parton of a particular flavour. Often, properties of the reconstructed jet are combined in a multivariate approach.

Two such techniques have been employed for the tagging of gluon and b jets:

The Quark-Gluon-Likelihood [125] takes into account the particle multiplicity, the jet shape and the fragmentation function. These three variables are combined into a multivariate discriminator. It assumes values between 0.0 for gluon jets and 1.0 for quark jets.

The Combined Secondary Vertex (CSV) Tagger [126] The most prominent feature of b jets is a secondary decay vertex. B quarks fragment into b hadrons. The decay of b hadrons is suppressed, resulting in a long lifetime with respect to their mass. This leads to an average decay length of a few centimetres, still within the volume of the CMS tracker. From the tracks of the decay products, the secondary vertex can be reconstructed. If such a secondary vertex is detected close to the jet vector, a high probability for a b jet is given. Other quantities such as the jet shape or the particle multiplicity are additionally included in this method.

The distribution of the two discriminators in simulation is shown in Figure 5.29. The distribution is shown for different jet flavours (determined with the physics definition).

Both tagging variables succeed in separating gluon respectively b jets from other flavours. Because of b jets being easily identifiable on the basis of their secondary decay vertices, the CSV tagger achieves high purity at high discriminator values. As jets from c hadrons show similar behaviour as b jets, such as secondary decay vertices from c hadron decays, c quarks also have higher discriminator values.

For the quark gluon tagger, the distinction is not quite as clear since gluon jets do not exhibit a prominent feature such as displaced vertices. Therefore, even at discriminator values close to zero, the fraction of gluon jets is only around two third. However, the discriminator succeeds in separating (light) quark and gluon jets, shown by the low gluon jet fraction at higher values.

The 2D-Tagging Method

The two discriminators can be combined for an improved estimation of the jet flavour and determination of the jet flavour response. This is done via the *2D-tagging method*.

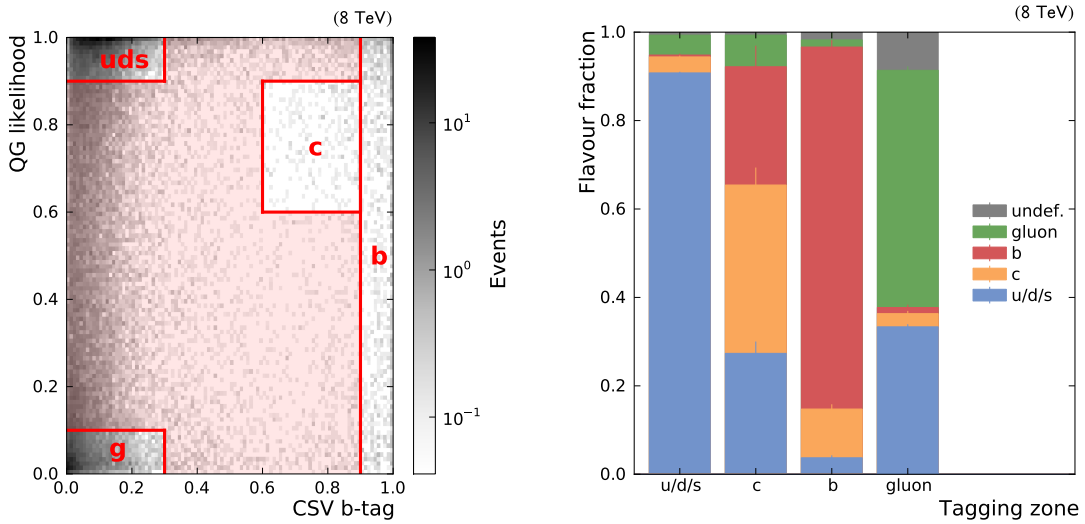
In the two-dimensional space spanned by the b and gluon tagging discriminators, four regions are defined with a high purity of light quark, gluon, c and b jets, respectively. In each region, the jet flavour composition (from simulation, using the physics definition) and the jet response (in data and simulation) are determined. Combining these results, the jet response per flavour can be determined in data and simulation. The jet response in data and simulation can then be compared to derive data-driven flavour corrections.

The distribution of events in the 2D-tagging plane and the definition of the four zones can be seen in Figure 5.30(a). Due to the much larger number of gluon or light quark jets in the Z+jet sample, the region of high values of the b tagging discriminator is only sparsely populated. This results in a higher statistical precision of the results from the light quark and gluon zones compared to the c or b zones.

The definition of the tagging zones has been optimized to decrease the statistical uncertainty on the final result, the jet response per flavour. This requires to optimize the zone definitions for purity while maintaining an acceptable event count per zone. In Figure 5.31, the jet flavour fractions determined from simulation are shown as a function of the two tagging discriminators. The tagging zone for each flavour is also drawn. Within these zones, the fraction of the respective flavour is particularly high.

The fractions of jet flavours for each tagging zone are also visualized in Figure 5.30(b). As could already be seen in Figure 5.31, the purity in the c zone is lowest, followed by the gluon zone. The highest purity can be found in the b and light quark zones.

The measurement of the MPF response for each zone can be seen in Figure 5.32(a). As expected from the results in simulation (see Figure 5.28(a)), the MPF response for c and b jets is considerably lower compared to light quark jets because of the missing energy from neutrinos. Gluon jets have slightly lower response caused by their broader fragmentation.



(a) Distribution of the tagging variables and the tagging zones

(b) Flavour (from simulation) composition for each tagging zone

Figure 5.30: Zones and Flavour composition in the tagging zones.

The results from the response measurement and the composition study now have to be combined to determine the response for each individual flavour. This is done by solving the following equation:

$$\begin{pmatrix} R_1 \\ R_2 \\ R_3 \\ R_4 \end{pmatrix} = \begin{pmatrix} f^1_{\mathrm{uds}} & f^1_{\mathrm{c}} & f^1_{\mathrm{b}} & f^1_{\mathrm{g}} \\ f^2_{\mathrm{uds}} & f^2_{\mathrm{c}} & f^2_{\mathrm{b}} & f^2_{\mathrm{g}} \\ f^3_{\mathrm{uds}} & f^3_{\mathrm{c}} & f^3_{\mathrm{b}} & f^3_{\mathrm{g}} \\ f^4_{\mathrm{uds}} & f^4_{\mathrm{c}} & f^4_{\mathrm{b}} & f^4_{\mathrm{g}} \end{pmatrix} \times \begin{pmatrix} R_{\mathrm{uds}} \\ R_{\mathrm{c}} \\ R_{\mathrm{b}} \\ R_{\mathrm{g}} \end{pmatrix} .$$

R_x is the response for each zone where x being 1/2/3/4 corresponds to the light quark, gluon, c and b zone, respectively. f^y_z is the fraction of flavour z in zone y. Solving this equation with the previously determined values yields the values for the response per flavour R_z.

The results for R_z are shown in Figure 5.32 for data and simulation. In addition to the result for the flavour from tagging, for the simulation the MPF response as a function of the flavour determined with the physics definition is also shown. For b and light quark jets, the result for the respective zones and for the flavour determined from tagging is similar because of the high purity within the tagging zone. For c jets, the relatively low purity and the small total event count in the respective tagging zone leads to a high statistical error on the final result and therefore low reliability. The results for flavour from tagging disagree with the results for flavour from simulation (physics definition), likely due to the limited efficiency of the tagging methods. For light quark and even more for b jets, differences in the jet response between data and simulation can be observed.

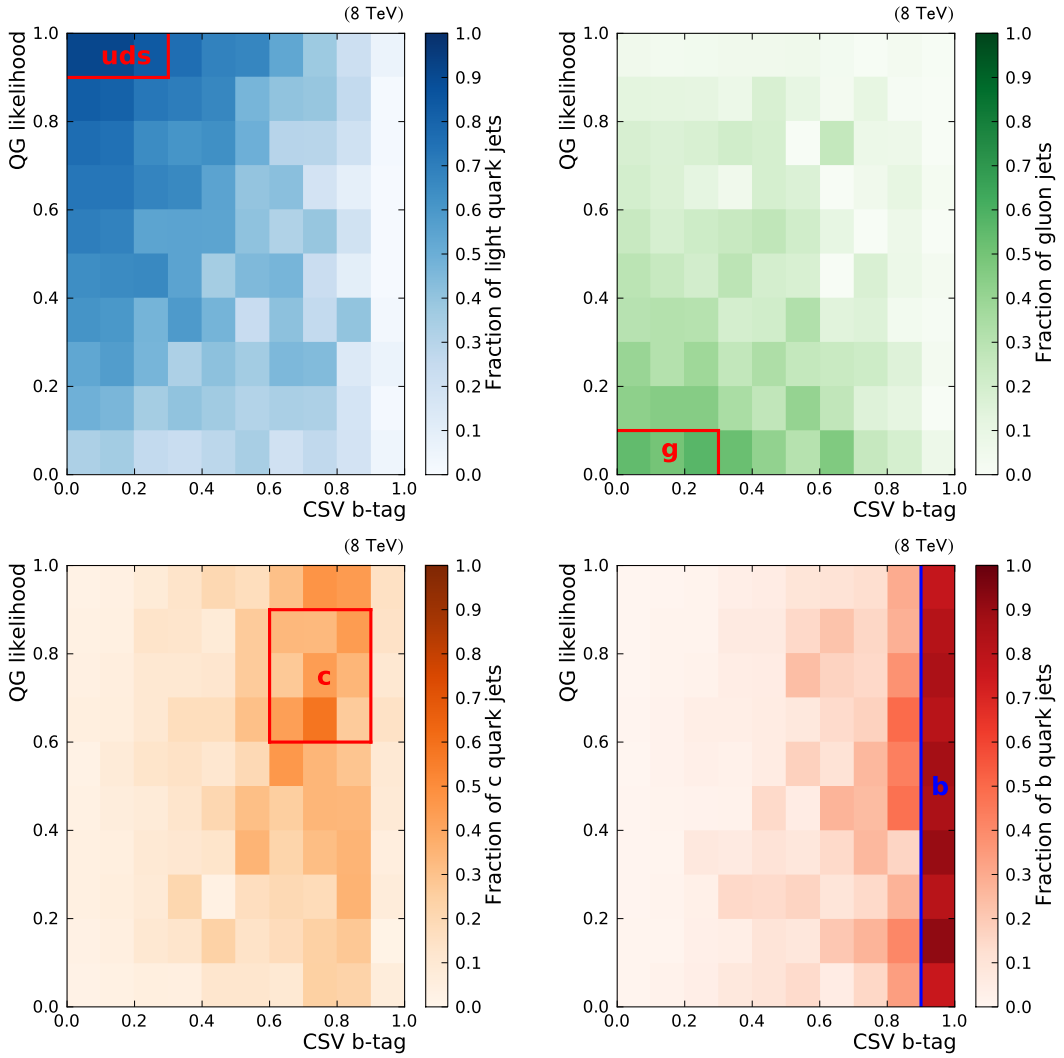
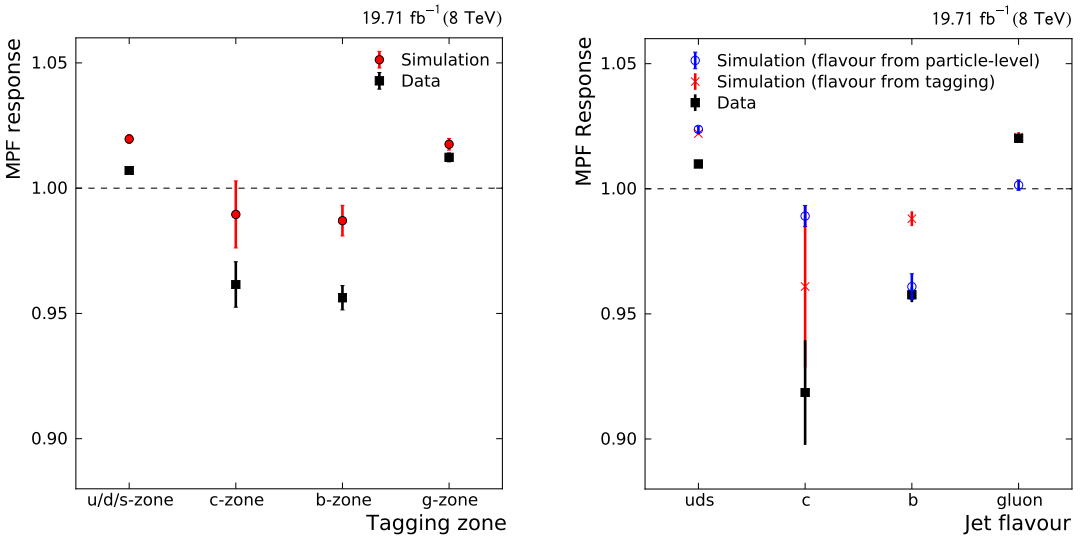


Figure 5.31: Flavour fractions as function of b-tag and quark-gluon-tag, also showing the zones used for 2D-tagging studies with boxes. The fractions and zones for light quark (upper left), gluon (upper right), c (lower left) and b (lower right) jets are shown. The zones definitions have been determined to achieve a high purity for the respective flavour.

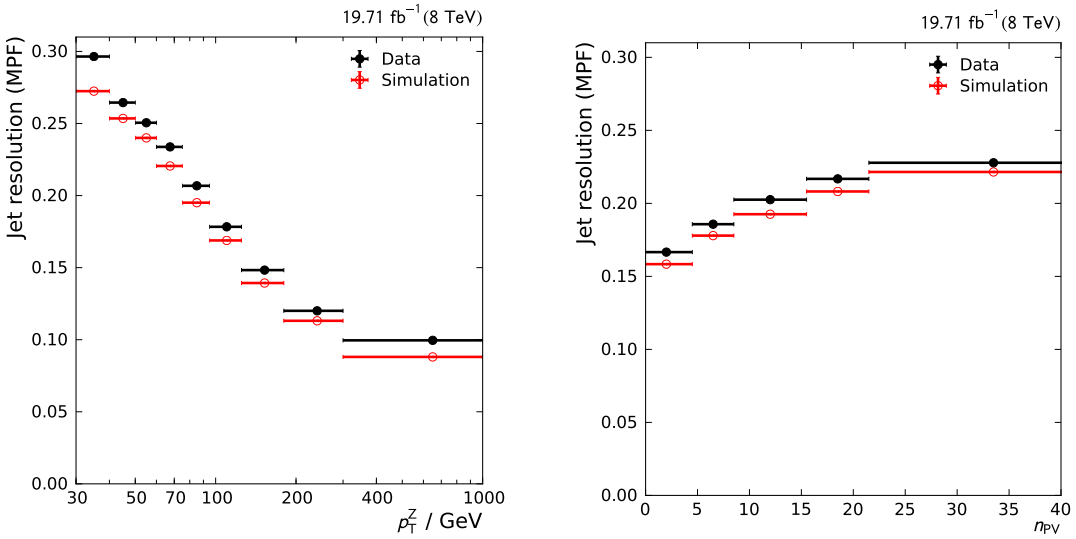


(a) MPF response for each tagging zone.

(b) MPF response for the individual flavours as a result of the 2D-tagging method.

Figure 5.32: MPF response for tagging zones and individual flavours (the result of the 2D-tagging method) in data and simulation. For the simulation, the results for the flavour from tagging and from the physics definition are shown.

The results obtained in this section provide extensive knowledge on flavour dependencies of the jet response. For the CMS Collaboration, these studies are important for a detailed estimation of flavour-induced JES uncertainties and dedicated flavour corrections.



(a) Resolution as a function of Z boson p_T

(b) Resolution as a function of N_{PV}

Figure 5.33: Jet p_T resolution in data and simulation, determined with the MPF method.

5.6 Jet p_T Resolution Measurement

For precision studies involving jets, not only the average jet response, but also the jet p_T resolution is important. Because of the steeply falling jet p_T spectrum in most physics analyses, a bad resolution can lead to an effective migration towards higher p_T bins and shift the measured distribution.

The jet p_T resolution corresponds to the width of the jet response distribution (which is shown in Figure 5.17). While the *mean* of the distribution is a measure for how well reconstructed jets approximate the true jet momenta on average, the *width* of the distribution shows how much the values are spread out around the mean. A small width, i.e. a small average deviation, corresponds to a good resolution.

The jet p_T resolution can be determined as the width of the MPF response distribution in bins of Z boson p_T, which represents the particle jet momentum in balanced events, see Figure 5.33(a). The corresponding plots for p_T balance are shown in Appendix A.2.1. The resolution worsens in the low p_T region. This is expected because of the bias from pileup, smearing or out-of-cone effects increasing at lower transverse momenta. The resolution is slightly worse in data compared to simulation, especially at lower p_T, possibly due to the mentioned effects not being perfectly simulated.

The jet response can also be studied for different pileup scenarios, represented by the number of reconstructed primary vertices. For a higher number of reconstructed vertices, i.e. higher pileup activity in the event, the resolution worsens. This is expected as pileup particles can add additional energy to the reconstructed jet momentum, thus biasing the measured response.

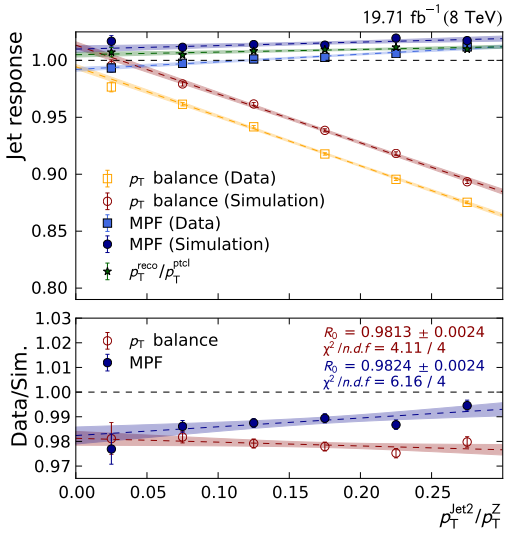


Figure 5.34: Extrapolation of the jet response to the ideal Z+jet topology, i.e. $p_T^{Jet2}/p_T^Z \to 0$. The p_T balance and MPF methods converge in the absence of additional jets. For both methods, the ratios between data and simulation also converge, enabling the extraction of the correction on the absolute JES.

Given the many systematic biases, good performance is achieved for most of the studied phase space. Only minor discrepancies between data and simulation are observed. Further results from jet resolution studies, e.g. as a function of jet pseudorapidity and using the p_T balance methods, can be found in Appendix A.2.1.

5.7 Determination of the Absolute Scale

A large number of studies on kinematic distributions, jet response and jet flavour have been performed. As the validity of the simulation and the employed methods has been verified, the absolute correction on the JES in data can be determined.

To this end, the results from the p_T balance and the MPF method have to be combined. As explained in Section 5.4, both methods are biased by systematic effects: the MPF response is influenced by pileup while the p_T balance method is affected by final-state radiation.

In order to correct for these effects, the results for each method have to be extrapolated to the ideal event topology with only one jet in the event, i.e. $p_T^{Jet2}/p_T^Z \to 0$. If additional jets, either from pileup or radiation, are suppressed, the two methods should converge and the absolute JES can be determined.

The result of this extrapolation is shown in Figure 5.34. In addition to the results for the p_T balance and MPF response in simulation and data, the simulated response is also shown as a function of p_T^{Jet2}/p_T^Z. A linear fit has been performed for all sets of values.

The values of the y-axis intercepts of the fitted lines are shown for the ratio between data and simulation for each method.

The p_T balance shows the expected behaviour of lower values for higher p_T of the second jet, i.e. if larger final-state radiation occurs and decreases the leading jet momentum. This behaviour is the same in data and simulation.

The fitted line on the MPF response also shows a slope, however it is much smaller compared to the results from p_T balance and points in the other direction: The values for the MPF response slightly decrease for lower values of $p_\mathrm{T}^\mathrm{Jet2}/p_\mathrm{T}^\mathrm{Z}$.

This trend is shared by the simulated response: The values are close to the results obtained with the MPF method. This behaviour is likely due to the effects from the underlying event: The simulation corrections are derived from a sample of dijet events, where the energy scale can be approximated by the jet p_T. For Z+jet events, the scale is higher at a given p_T as the mass of the Z boson has to be taken into account. The different scales result in a higher contribution from the underlying event in Z+jet events, especially at lower p_T, where the effect from the Z boson mass has stronger impact. As the average p_T of the leading jet decreases with $p_\mathrm{T}^\mathrm{Jet2}/p_\mathrm{T}^\mathrm{Z}$, this effect leads to the observed slope.

Both methods converge towards lower values of $p_\mathrm{T}^\mathrm{Jet2}/p_\mathrm{T}^\mathrm{Z}$. This shows that similar results are obtained with both methods if the bias from final-state radiation is corrected for. Minor remaining disagreements between the methods arise from residual detector miscalibration especially in the endcap and forward region, which affect the MPF response (through $E_\mathrm{T}^\mathrm{miss}$) but not p_T balance.

For the ratio between data and simulation, both methods show a small dependence on $p_\mathrm{T}^\mathrm{Jet2}/p_\mathrm{T}^\mathrm{Z}$. This can be explained by an imperfect modelling of final-state radiation in the simulation, which affects p_T balance, and remaining detector miscalibration, affecting the MPF response.

However, at $p_\mathrm{T}^\mathrm{Jet2}/p_\mathrm{T}^\mathrm{Z} = 0$ the ratio between data and simulation agrees between the two methods. From the values of the y-axis intercept, the jet response can be determined with high precision as $R = 0.982 \pm 0.003$. The correction factor on the absolute scale can be derived as the inverse of this value.

The extrapolation study proves that the two methods can be combined and any remaining biases do not affect the final result. To increase the precision even further, the correction on the absolute scale is determined by combining the results from different analyses.

5.8 Combination of Channels

The final correction on the absolute scale is determined by combining the results from Z($\rightarrow \mu^+\mu^-$)+jet, Z($\rightarrow \mathrm{e}^+\mathrm{e}^-$)+jet, γ+jet and multijet studies within their respective kinematic coverage. This combination increases the precision and accuracy of the correction. It is performed by a so-called global fit.

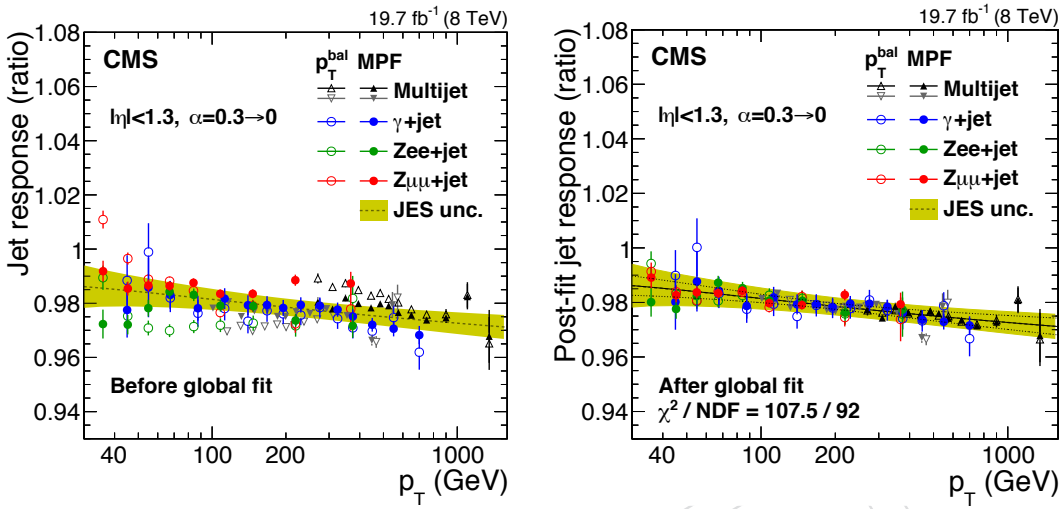


Figure 5.35: Response measurements in the Z($\to \mu^+\mu^-$)+jet, Z($\to e^+e^-$)+jet, γ+jet and multijet channels (left). A global fit is performed and the data points are shifted by the values for the nuisance parameters of the fit (right). As a result, all channels agree within uncertainties [109].

The results for each channel are corrected for final-state radiation and extrapolated to the ideal event topology of :

$$\frac{p_{\mathrm{T}}^{\mathrm{rad}}}{p_{\mathrm{T}}^{\mathrm{ref}}} \to 0 \tag{5.18}$$

where $p_{\mathrm{T}}^{\mathrm{ref}}$ is the p_{T} of the respective reference object, i.e. Z boson, photon or, in the case of multijet balancing, the jet recoil system. $p_{\mathrm{T}}^{\mathrm{rad}}$ refers to radiation p_{T}, i.e. the p_{T} of the second-leading jet in γ/Z+jet analyses or for multijet balancing to the highest-energetic jet which is not part of the balancing system.

The extrapolation is performed in bins of p_{T} of the reference object. Nuisance parameters are added to account for the uncertainty in lepton and photon scale, initial- and final-state radiation corrections, pileup and EM footprint. The fit is performed with two degrees of freedom for the absolute scale and the p_{T} dependence. The full details on the fitting procedure can be found in [109].

The result of the fit is shown in Figure 5.35. After the nuisance parameters have been determined by the fitting procedure, the data points are shifted by the respective values. As a result, the measurements from all channels agree within the uncertainty band. This result demonstrates the impressive level of detector calibration and understanding achieved by the CMS Collaboration.

5.9 Jet Energy Uncertainties

The total uncertainty on the jet energy scale is determined by combining several individual uncertainty sources. These belong to five major categories:

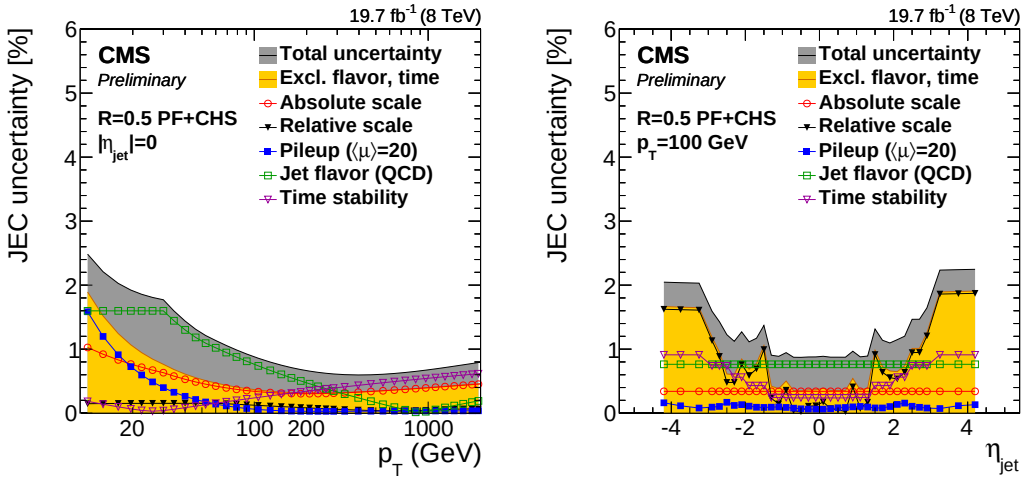


Figure 5.36: Jet energy uncertainty sources as a function of jet p_T for central jets (left) and as a function of $|\eta|$ for jets with $p_T = 100\,\text{GeV}$ (right). The uncertainties are largest in the low-p_T region, where pileup and flavour effects increase. In the endcap and forward regions, the uncertainty in the relative scale becomes dominant. In the central detector at medium to high p_T, the uncertainty is below 1% [109].

Pileup This uncertainty is determined by studying the difference between the measured offset and the true offset in simulation. The effects of varying ρ by one sigma are also taken into account. The pileup uncertainty is dominant at low p_T, as most pileup interactions produce jets with only low transverse momenta.

Relative The uncertainty on the relative scale calibration is mainly caused by the finite jet p_T resolution and the uncertainty on the modelling of soft radiation.

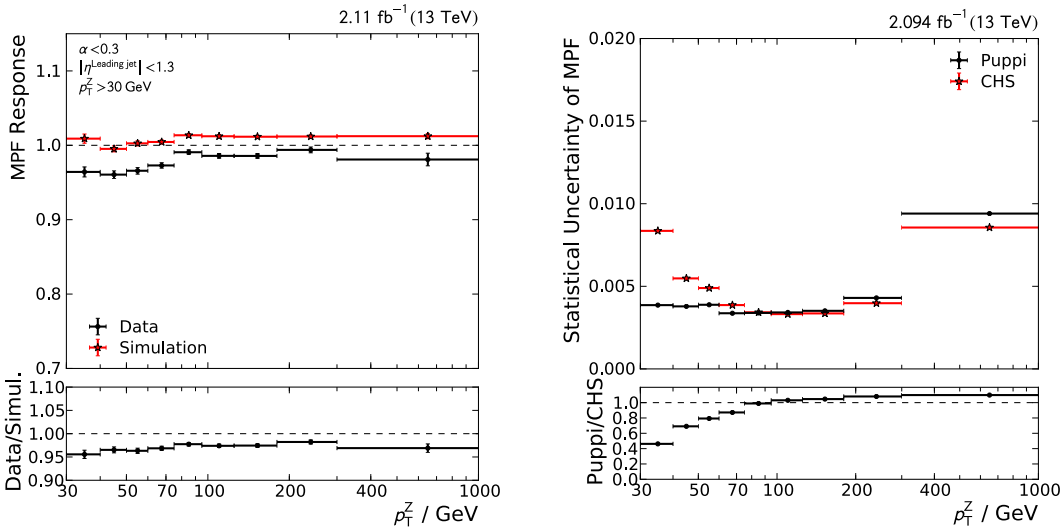
Absolute The absolute scale uncertainty incorporates mainly the uncertainty on the p_T dependence of the global fit and the single pion response.

Flavour This uncertainty is determined from differences between Pythia and Herwig. It is largest for gluon jets due to the different treatment of the hadronization process.

Time dependence Because detector conditions change over time (especially in the endcap region, see Figure 5.24), the JES can differ between detector run periods.

A more complete description of the uncertainty sources, their magnitude and correlations can be found in [109]. The values for the individual uncertainty contributions including correlations between different regions in jet p_T and η are centrally provided by the CMS Collaboration for all physics analyses. The uncertainties are visualized in Figure 5.36.

If the entire data collected by the CMS detector at $\sqrt{s} = 8\,\text{TeV}$ is used in an analysis (as is the standard for most physics analyses), the time dependence uncertainty can be



(a) MPF response as a function of p_T^Z in data and simulation [96].

(b) Statistical uncertainty of calibrating PF+CHS and PF+PUPPI jets [85].

Figure 5.37: Z($\rightarrow \mu^+\mu^-$)+jet calibration results at $\sqrt{s} = 13$ TeV. Left: For the MPF response as a function of p_T^Z, a small difference between data and simulation can be observed. Right: Comparison of the statistical calibration uncertainties of PF+CHS and PF+PUPPI [72] jets. At low p_T, the pileup mitigation of PUPPI improves the precision by a factor of two.

neglected. If the jet flavour composition of the used data sample matches the Z+jet flavour composition (for which the absolute calibration is derived) and jet flavour corrections are applied, the flavour uncertainty can also be neglected. The total uncertainty can then be as small as 0.35% in the central detector region for a jet with $p_T = 100$ GeV.

5.10 Outlook on Z+Jet Calibration at $\sqrt{s} = 13$ TeV

In 2015, Run 2 of the LHC has begun. The increased centre-of-mass energy of 13 TeV and doubled bunch crossing frequency pose even higher experimental challenges. In addition, the CMS detector needs to be recalibrated after upgrade and maintenance operations have been performed.

The Z($\rightarrow \mu^+\mu^-$)+jet calibration is continued at KIT. In addition to existing methods, new techniques, e.g. for pileup mitigation, are explored. First calibration studies are detailed in [85, 96, 117, 118]. A selection of the obtained results are shown in Figure 5.37.

5.11 Summary

The CMS Collaboration has demonstrated an outstanding performance in jet calibration, with a JES uncertainty as small as 0.35%. The continuous calibration efforts have considerably decreased the JES uncertainty for the data at $\sqrt{s} = 8\,\text{TeV}$, see Figure 5.38. Thereby, the experimental uncertainty for many physics analyses has been significantly reduced. New techniques have been explored to increase experimental precision and tackle the challenges of physics at the terascale.

The analysis of Z($\rightarrow \mu^+\mu^-$)+jet events presented in this thesis have made a decisive contribution to these results. Many independent validation studies, e.g. of the kinematic distributions of the involved physics objects or the Particle-Flow event reconstruction, were performed. The predictions from simulation are mostly confirmed. The jet response has been measured and compared to simulation, using two complementary methods. The behaviour of the jet response for different phase space regions as well as the relative merits of the two methods have been studied. The precision on the JES was greatly increased with detailed studies of jet flavour effects, using novel approaches and techniques such as multivariate tagging methods. The correction on the absolute JES has been derived from a combination with other channels by means of a global fit.

The pioneered techniques and the obtained results constitute an excellent foundation for Run 2 of the LHC. For the data-driven jet calibration, the Z($\rightarrow \mu^+\mu^-$)+jet analysis remains the pillar of precision.

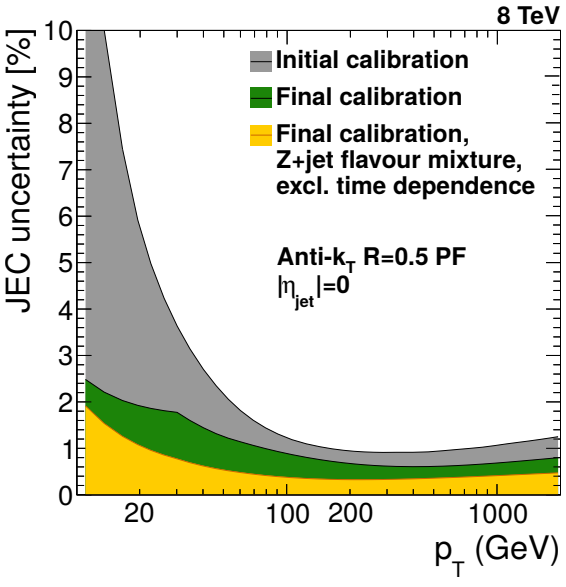


Figure 5.38: Evolution of jet energy uncertainties. The initial calibration at the beginning of the data taking period (grey), performed with a data sample corresponding to $1.9\,\text{fb}^{-1}$, is compared to the final calibration (green). The uncertainties are reduced over the entire p_T range.
If the jet flavour composition in a sample corresponds to the Z+jet flavour mixture and the full $19.7\,\text{fb}^{-1}$ of data are used, the uncertainty is even lower (yellow).
The uncertainties are reduced mainly because of the increase of data for calibration, the sophistication of calibration techniques. and the improved estimation of flavour uncertainties. Adapted from [109, 111].

6 Constraining PDFs with $Z(\rightarrow e^+e^-)$+Jet Events

The structure of the proton is described by the parton distribution functions (PDFs), see Section 2.1.2. In perturbative QCD, they are an essential ingredient for cross section calculations in proton collisions. To reduce the uncertainty in the calculated cross section, the PDFs need to be known as precisely as possible.

However, the PDFs cannot be calculated from first principles but have to be experimentally determined. They are not directly measurable but can only be determined via their correlation with observables such as cross sections in proton scattering experiments.

This correlation was for example exploited in deep-inelastic scattering experiments at the HERA collider [9, 127] to obtain considerable knowledge about the PDFs. Dedicated collaborations such as NNPDF [11], CTEQ [13] or MMHT [14] further constrain the PDFs using the results of several experiments. To further reduce the PDF uncertainties, additional measurements have to be very precise or study unexplored regions in x or Q.

At the LHC, the unprecedented centre-of-mass energy of proton-proton collisions allows for processes with higher Q^2 compared to previous experiments, see Figure 2.4. The CMS detector provides the kinematic coverage to reconstruct processes involving a wide range in x of the interacting partons. For the calculation of LHC cross sections, determining the PDFs in the same energy regime would increase the reliability of the theory calculations. However, most analyses at the LHC suffer from large systematic uncertainties, e.g. from pileup, background processes or from the difficult reconstruction of objects such as jets or neutrinos.

The measurement of the $Z(\rightarrow e^+e^-)$+jet cross section avoids some of these problems: The low background contributions and the high precision in electron[1] reconstruction result in small systematic uncertainties. At the LHC, Z+jet events are produced at a rate sufficient for precision measurements. Therefore, Z+jet studies are promising in terms of PDF constraints.

The following sections describe the extraction of PDF constraints from a measurement of the $Z(\rightarrow e^+e^-)$+jet cross section double-differentially in Z boson p_T and rapidity. Theory predictions are computed with Monte Carlo simulations at next-to-leading-order accuracy in perturbative QCD. For the analysis of detector data, the Z boson reconstruction requires the careful calibration of the electron energy scale and the determination of reconstruction, identification and trigger efficiencies. The contributions from background processes are estimated and the measured distributions are unfolded to correct for reconstruction biases. The dominant systematic uncertainties are determined.

[1] In this chapter, 'electron' refers to both electrons and positrons.

The results of measurement and simulation are compared in order to determine the PDFs. A combined fit of the PDF parameters to HERA I+II deep-inelastic scattering and CMS Z(→ e⁺e⁻)+jet data is performed. The impact of the Z(→ e⁺e⁻)+jet measurement on the PDFs and their uncertainties is determined and discussed. The technical workflow is shown in Figure 4.2. The analysis is also detailed in [128].

6.1 Simulation of Z(→ e⁺e⁻)+Jet Events

The Z(→ e⁺e⁻)+jet differential cross section has been calculated with SHERPA [90] at next-to-leading-order accuracy at $\sqrt{s} = 8\,\text{TeV}$. The virtual corrections were provided by BLACKHAT [129]. The configuration file used is shown in Appendix A.3.1. In total, 6×10^9 events have been simulated. As described in Section 4.3, the FASTNLO [98] tables necessary for PDF fits were created from the generated events via RIVET [130] and MCGRID [100]. The factorization and renormalization scales for each event have were set to

$$\mu_f = \mu_r = \sqrt{m_\text{Z}^2 + p_\text{T,Z}^2}\,.$$

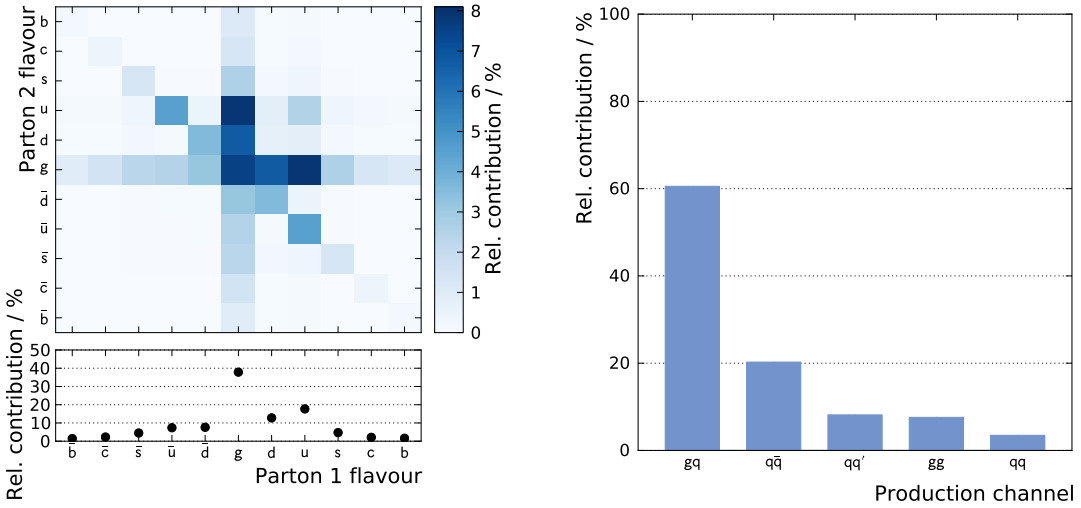
Figure 6.1(a) shows the flavour distribution of the initial partons. As expected, the distribution is symmetric. For each of the two interacting protons, the gluon makes up the dominant contribution, followed by quarks from light to heavy. The u and d quark contributions show an excess (which is twice as high for u quarks compared to d quarks) with respect to the corresponding antiquarks. These results are expected as the distribution of flavours corresponds to the composition of the proton, which is expressed by the PDFs as shown in Figure 2.5.

The Z+jet production channels are determined by sorting the individual contributions for each flavour combination into five categories: gq, gg, qq' (two quarks of different flavour), qq (two quarks of the same flavour) and $q\bar{q}$. Some of the corresponding Feynman diagrams are shown in Figure 2.2. The gq channel is dominant, see Figure 6.1(b). The $q\bar{q}$ channel, which corresponds to the leading-order diagram of Z boson production (see Figure 2.7) with additional gluon radiation, makes up the second highest contribution. Adding the gq and gg categories together, two-thirds of events involve at least one gluon. Quarks are involved in 90% of all events (every production channel except gg), distributed among all flavours with a preference towards light and valence quarks.

Figure 6.2 shows the predicted cross section as a function of Z boson absolute rapidity and p_T. To validate the FASTNLO interpolation procedure, the evaluation of the FASTNLO table with the same PDF as used for the original SHERPA calculation is studied. The differences between the SHERPA prediction and FASTNLO are at sub-permille level, illustrating the high accuracy of the FASTNLO interpolation method.

6.1.1 Sensitivity of the Cross Section to PDFs

The determination of the PDFs relies on their correlation with observable distributions. This dependence can be quantified by studying the effect of varying the PDFs on the Z(→ e⁺e⁻)+jet cross section.



(a) Flavours of the initial partons in Z+jet events

(b) Z+jet production channels

Figure 6.1: Initial parton flavour composition in Z+jet events. The dominant contributions come from gluons and light quarks, in particular valence quarks.

For this purpose, a PDF set from the NNPDF collaboration is used. This set consists of N PDF replicas which is each obtained by randomly varying (according to the respective uncertainties) the experimental results the PDF determination is based upon. The effect of variations of the PDF for flavour f on the cross section σ_{Z} depending on a variable a (e.g. p_{T} or rapidity of the Z boson) is quantified via the correlation coefficient

$$\rho_f(x, a, Q) = \frac{N}{N-1} \frac{\langle \sigma_{\mathrm{Z}}(a)_i \cdot xf(x, Q^2)_i \rangle - \langle \sigma_{\mathrm{Z}}(a)_i \rangle \cdot \langle xf(x, Q^2)_i \rangle}{\Delta_{\sigma_{\mathrm{Z}}(a)} \Delta_{xf(x,Q^2)}} \,. \tag{6.1}$$

The angular brackets indicate averaging over all PDF replicas, with the PDF or cross section for an individual replica denoted by the index i. The standard deviation of the cross section or the PDF over all replicas is denoted by Δ. To correct for the bias arising from the mean and standard deviation determined from the same sample, Bessel's correction $N/(N-1)$ is applied.

The correlation coefficient ρ assumes values between +1 and -1, corresponding to high correlation and anti-correlation, respectively. PDF constraints can be particularly expected in phase space regions with a high correlation coefficient, i.e. where the cross section strongly depends on the PDF.

The obtained correlation values are shown in Figure 6.3 for different flavours as a function of x and the p_{T} or the rapidity of the Z boson. The PDFs are evaluated at $Q = m_{\mathrm{Z}}$. The sea quark PDF is defined as the double of the sum of the PDFs for $\bar{\mathrm{d}}$, $\bar{\mathrm{u}}$ and $\bar{\mathrm{s}}$ quarks, assuming the sea quark distributions are equal for the corresponding flavours of quarks and antiquarks.

For the correlation of the gluon PDF as a function of the Z boson p_{T} (upper left), the correlation for higher values of x increases with p_{T}. Complementarily, the correlation

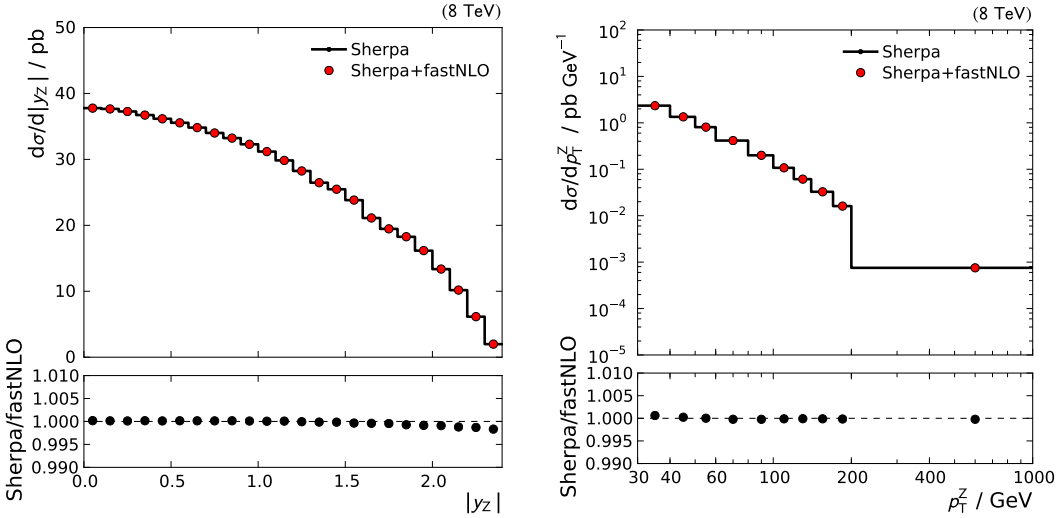


Figure 6.2: Comparison of the original SHERPA +BLACKHAT prediction with the FASTNLO interpolation, needed for PDF fits. The differences between SHERPA and FASTNLO are at sub-permille level, illustrating the high accuracy of the FASTNLO interpolation procedure.

for medium-to-lower x values increases with rapidity (upper right). This corresponds to the kinematic relations of the collision: High transverse momenta of the outgoing objects require high momentum fractions x of both interacting partons. In contrast, a high rapidity necessitates a lower x of one of the partons compared to the other.

For the u and d valence quark (second and third row), the correlations between PDF and σ_Z is not quite as high as for gluons. Minor correlation is observed in regions of low and high x. For the sea quark PDF (bottom row), a mild correlation can be observed at medium x at lower and medium p_T and over the entire $|y_Z|$ range.

Correlations in the high- and low-x regions correspond to high p_T and high absolute rapidity of the Z boson, respectively. However, these phase space regions are statistically limited. In contrast, the regions of low p_T and low to medium absolute rapidity offer the highest statistical precision. For these reasons, PDF constraints can be expected over a wide range in x.

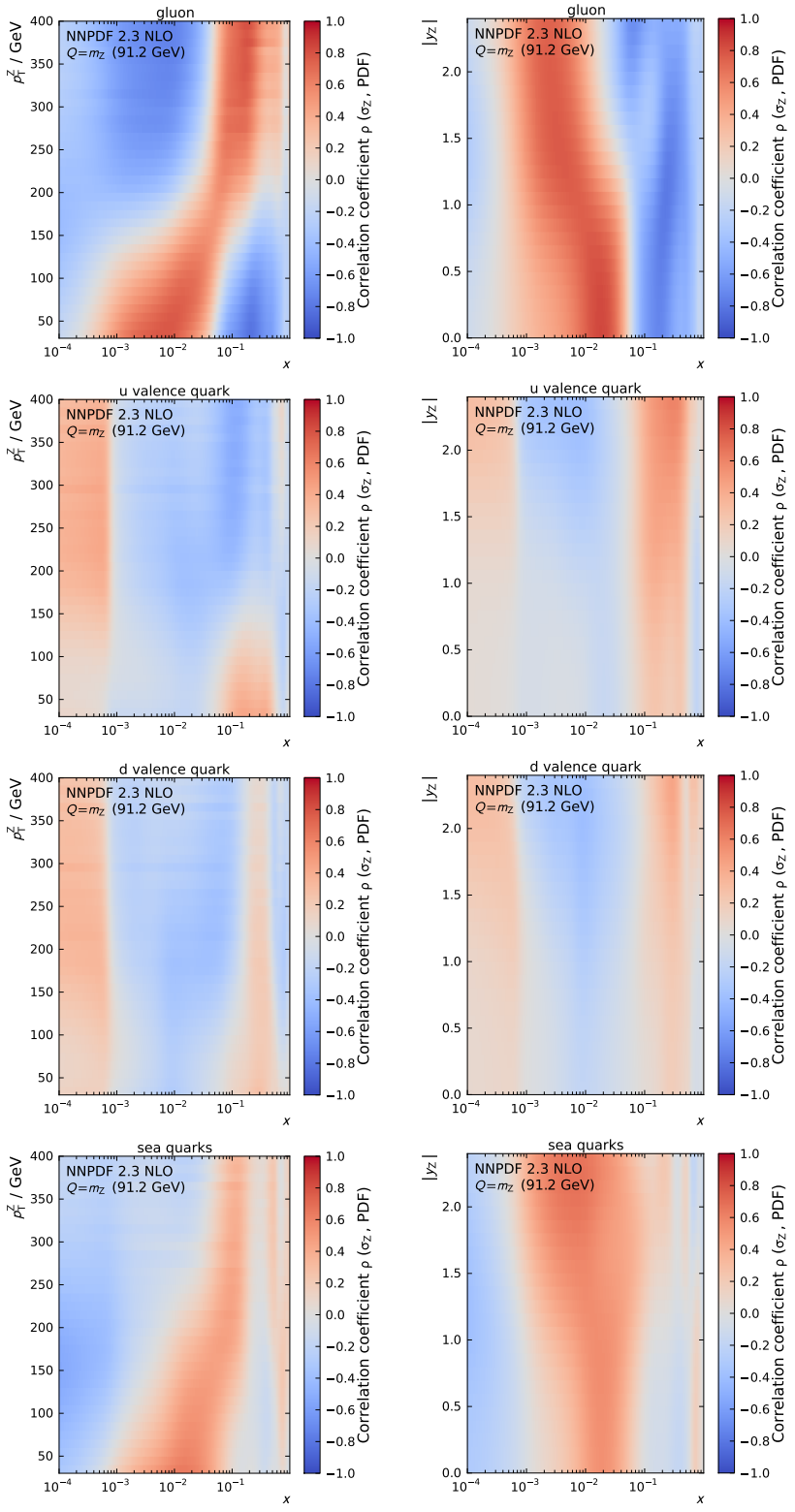


Figure 6.3: Correlation between Z+jet cross section and PDFs. The correlation coefficient is shown for different flavours (rows) as a function of x and p_{T} (left column) or rapidity (right column) of the Z boson.

6.2 Measurement of Z(→ e⁺e⁻)+Jet Cross Section

The measurement is performed double-differentially in Z boson p_T and absolute rapidity. The binning in p_T is 30, 40, 50, 60, 80, 100, 120, 140, 170, 200 and 1000 in units of GeV. For the absolute rapidity, the binning is 0, 0.4, 0.8, 1.2, 1.6, 2.0 and 2.4 for the double-differential measurement. For the projection onto only the y_Z axis, 24 bins in rapidity are used from 0.0 to 2.4 with a constant bin width of 0.1.

The following sections describe the calibration of the electron energy scale, the correction for reconstruction inefficiencies and the event selection. The background contributions are estimated and the measured distributions are unfolded.

6.2.1 Electron Momentum Corrections

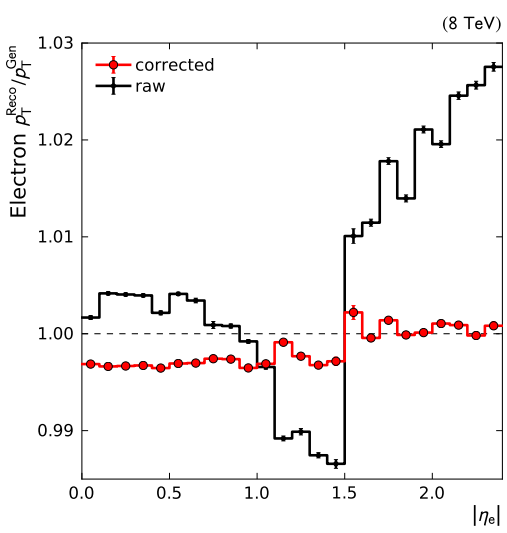
The reconstruction of electrons with the CMS detector is described in Section 3.2.2. Energy and momentum of an electron are determined by combining information from the ECAL (calorimeter deposits) and the tracker (track curvature). Unfortunately, the electron reconstruction is biased by effects such as energy losses due to bremsstrahlung, energy leakage out of the ECAL or additional energy from pileup.

The ECAL response is calibrated and the radiated energy from bremsstrahlung photons is taken into account for track reconstruction. In order to further increase the precision of the energy calibration, additional corrections are applied to the reconstructed electron. A multivariate approach is used, combining parameters of the ECAL supercluster, the track and the reconstructed electron, e.g. the shape or size of the supercluster. The procedure is detailed in [57].

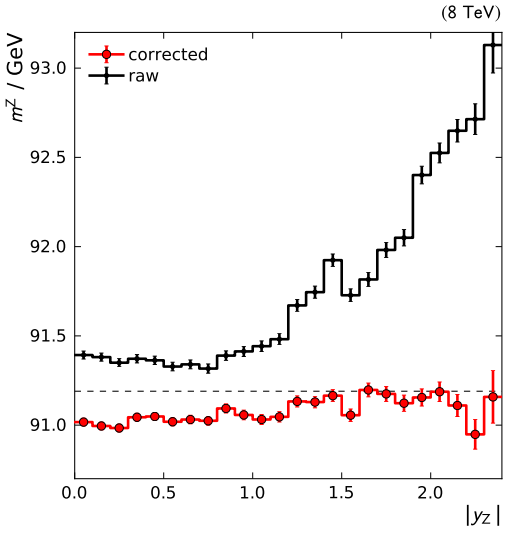
The effect of the corrections is visualized in Figure 6.4(a). In a sample of simulated Z boson decays, particle-level electrons are matched to reconstructed electrons by requiring $\Delta R < 0.3$. The p_T-ratio of a reconstructed electron and the matching particle-level electron is studied as a function of electron absolute pseudorapidity. Without corrections, a large dependence of the reconstruction performance on the pseudorapidity region is visible. The reconstructed electron underestimates the particle-level p_T in the ECAL barrel section close to the transition region ($1.0 < |\eta| < 1.5$). At higher absolute pseudorapidities, the reconstruction overestimates the electron p_T by almost three percent. After applying the momentum corrections, the pseudorapidity-dependent bias is greatly reduced.

The invariant mass of the reconstructed Z boson is studied as a function of the Z boson rapidity, see Figure 6.4(b). For uncorrected electrons, the variations in the electron energy scale lead to a significant rapidity-dependence of the reconstructed Z boson mass. After the corrections have been applied, this dependence has largely vanished. The remaining bias is absorbed into the electron scale uncertainty.

Both studies show the success of the electron momentum corrections in further improving the experimental precision by decreasing reconstruction biases.



(a) p_T^{reco}/p_T^{ptcl} of uncorrected ("raw") and corrected electrons as a function of absolute electron pseudorapidity.



(b) Invariant mass of the reconstructed Z boson as a function of its rapidity.

Figure 6.4: Effects of electron momentum corrections. The corrections greatly reduce the electron response differences between the different detector regions (left). Subsequently, the average Z boson invariant mass as a function of rapidity is stabilized (right).

6.2.2 Electron Reconstruction, Trigger and Identification Efficiencies

As not all $Z \to e^+e^-$ events are properly reconstructed, the measured event rate is lower than the actual one. The reasons are imperfections in reconstruction, triggering and identification of electrons. These effects have to be quantified and corrected for.

The efficiencies, i.e. the probabilities of electrons to be properly reconstructed and selected, are first determined from the detector simulation and then corrected with a data-driven method known as *tag and probe*. This technique determines the relative efficiency of a tight selection with respect to a loose selection. Electrons are classified into the (non-exclusive) categories of *probe* if they pass the loose selection and *tag* if they pass the tight selection. Pairs of a tag and a probe electron with an invariant mass close to the Z boson resonance (60 GeV to 120 GeV) are selected. In this region, the overwhelming majority of dielectron pairs originates from Z boson decays. The background contributions are small and subtracted via a fit to the mass distribution.

Events are classified depending on whether the probe electron fails or passes the tight selection requirement. Based on the numbers of events with failing and passing probes, the relative efficiency of the tight with respect to the loose selection can be determined. The efficiencies are computed for three steps: reconstruction of electron candidates (track-supercluster matching), identification and triggering.

The efficiency usually has a turn-on behaviour, i.e. a sharp increase at lower p_T, where reconstruction is more difficult, and is constant above 50 GeV. Therefore, a finer binning

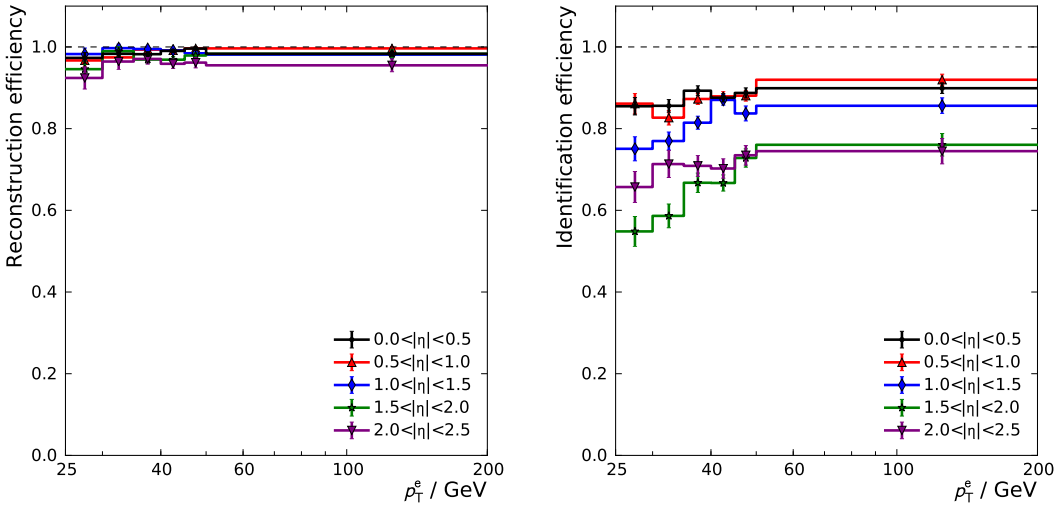


Figure 6.5: Efficiencies of electron reconstruction (left) and identification (right). The efficiencies are shown as a function of p_T for different regions of electron absolute pseudorapidity. Efficiencies decrease for higher absolute pseudorapidity and lower p_T of the electron. Scale factors are derived as the inverse of the efficiency for each region in p_T and $|y|$.

is chosen at lower p_T. The binning in $|\eta|$ corresponds to the ECAL geometry, with the inner and outer parts of the ECAL barrel and endcap section, respectively, and the barrel-endcap gap at $1.442 < |\eta| < 1.566$. Systematic and statistical uncertainties are properly taken into account. More details on the method and the implementation can be found in [131].

The reconstruction and identification efficiencies are shown in Figure 6.5. For the identification efficiency, the expected turn-on at lower p_T is visible. For $p_T < 25\,\text{GeV}$, the efficiency is expected to drop further, however electrons in this phase space region are not included in this analysis. The lower efficiency in the ECAL endcap ($|\eta| > 1.5$) is due to the lower energy resolution and less uniform detector geometry. Given the tight identification criteria, the trigger efficiency is >99% in all regions. The corresponding values are shown in Appendix A.2.2.

The scale factors c are determined as the inverse of the efficiency in bins of electron $|\eta|$ and p_T for all three steps. They are applied for each of the two decay electron candidates per event as a weight:

$$w_{\text{event}} = c_{1,\text{reco}} \cdot c_{1,\text{id}} \cdot c_{1,\text{trig}} \cdot c_{2,\text{reco}} \cdot c_{2,\text{id}} \cdot c_{2,\text{trig}} \,. \tag{6.2}$$

6.2.3 Event Selection

As a first requirement, the full functionality of the CMS detector has to be ensured for the recorded data. A list of certified run periods is centrally provided by the CMS Collaboration. Events which were not recorded during these periods are excluded. The

certified data at $\sqrt{s} = 8\,\text{TeV}$ correspond to $19.71\,\text{fb}^{-1}$ of integrated luminosity. A software trigger which selects events containing one electron with $p_\text{T} > 17\,\text{GeV}$ and a second one with $p_\text{T} > 8\,\text{GeV}$ is used. To lower the misidentification rate, electron candidates are subjected to several quality criteria. The Z boson can then be reconstructed from the two decay electrons, i.e. by addition of their four-vectors. The full technical details on data certification, trigger names, identification and isolation requirements are given in Appendix A.1.

The following kinematic selection is applied:

- Electron reconstruction is difficult at low p_T, leading to low efficiency and possible differences between data and simulation. To avoid these effects, a lower threshold is applied to the p_T of each electron:

$$p_{\text{T,e}} > 25\,\text{GeV}\,. \tag{6.3}$$

- For similar reasons, the detector regions where electron reconstruction is more challenging are excluded. This concerns the transition between ECAL endcap and barrel sections and the border of the tracker. The following limits are applied to the pseudorapidity of each electron:

$$|\eta_\text{e}| < 1.442 \vee |\eta_\text{e}| > 1.566\,, \tag{6.4}$$

$$|\eta_\text{e}| < 2.4\,. \tag{6.5}$$

- A Z boson candidate is found if two oppositely charged electrons in an event fulfil the above criteria. The invariant mass of the Z boson candidate, i.e. of the dielectron system, has to be close to the nominal Z boson mass:

$$|m_\text{ee} - 91.19\,\text{GeV}| < 10\,\text{GeV}\,. \tag{6.6}$$

- The kinematic range is restricted to a region where a fixed-order calculation of the Z(→ e⁺e⁻)+jet cross section is reliable and low-p_T effects such as soft parton emissions become negligible. Therefore, a lower limit on the Z boson p_T is set:

$$p_{\text{T,Z}} > 30\,\text{GeV}\,. \tag{6.7}$$

To avoid the large uncertainties in the jet energy scale, no requirements are imposed on jets.

6.2.4 Estimation of Background Contributions

Two isolated electrons in the final state can originate from a number of physics processes other than Z boson decay in Z+jet events. The contributions from these processes have to be determined and subtracted from the measured distributions. For each process, a sample of simulated events is subjected to the same event selection procedure as applied

to data. The resulting event count is normalized according to the process cross section and the integrated luminosity of the data to obtain the background estimation.

The considered background processes can be classified into the following categories according to their final state:

Two electrons from Z boson decay, e.g. from electroweak diboson processes such as WZ and ZZ production.

Two electrons from other decays, e.g. from $t\bar{t}$, tW, WW and $Z \to \tau^+\tau^-$ processes which can lead to a final state with two electrons.

At least one jet being misidentified as electron. This category can be further divided depending on the number of misidentified electrons:

dijet processes where both jets are falsely reconstructed as electrons. The probability for this case is rather low, however the large cross section of dijet production could lead to substantial background contributions.

W+jet production. In case of a W boson decaying into an electron, an electron-neutrino and an additional jet which is reconstructed as an electron, this process is identified as a dielectron final state.

Exemplary Feynman diagrams for some of these processes are shown in Figure 6.6. The cross sections for all considered background processes are reported in Table 6.1. Additionally, the number of events and CMS-internal names of the simulated samples are given in Appendix A.1.

Figure 6.7(a) shows the distribution of background events as a function of the invariant mass of the dielectron pair. Overall, WZ and ZZ make up the largest background contribution, followed by $t\bar{t}$ production. In WZ and ZZ samples, where the two electrons originate from a Z boson decay, the Z boson mass peak is clearly visible. Even though the cross section of diboson production is comparatively low, the dielectron final state is indistinguishable from Z+jet events. All other studied background processes yield only minor contributions, mainly because of the tight electron selection and therefore low rate of jets being misidentified as electrons. The background contributions depending on the number of jets in the event, the rapidity or p_T of the Z boson as well as the relative contributions of different processes to the total background are shown in Figure A.12. The background contributions for the double-differential measurement are shown in Figure A.13.

The sum of the studied background contributions is compared to the expected number of signal events, see Figure 6.7(b). As can be seen from the background-signal-ratio, the relative background is lowest close to the peak of the Z boson mass where the signal process is dominant. Correspondingly, the relative background contribution slightly increases further away from the Z boson peak. The background/signal ratio as a function of Z boson p_T or rapidity or for the double-differential measurement is shown in Figure A.14.

Overall, the relative background is at the level of only few percent, mainly due to the large cross section of the signal process and the low misidentification rate of electrons. To correct the data for background contributions, the estimated number of background events is subtracted from the measured number of events in the respective bins.

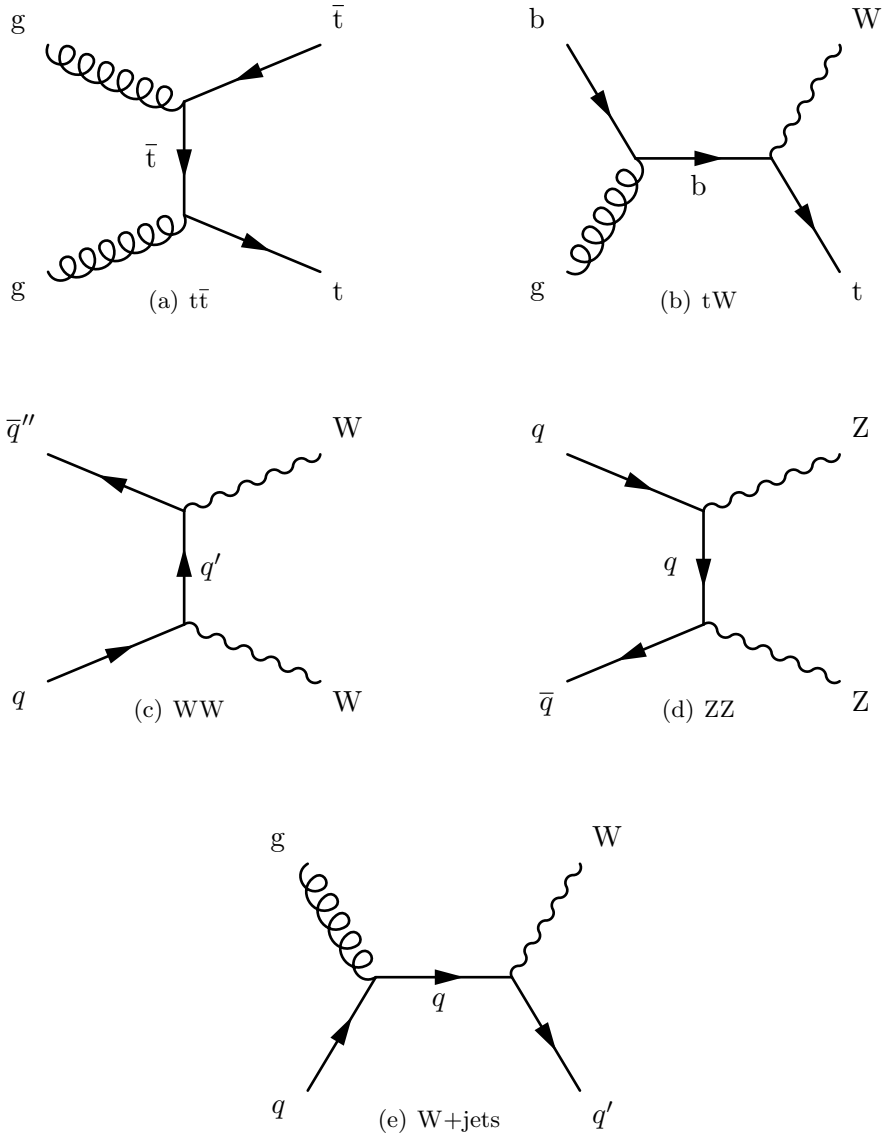
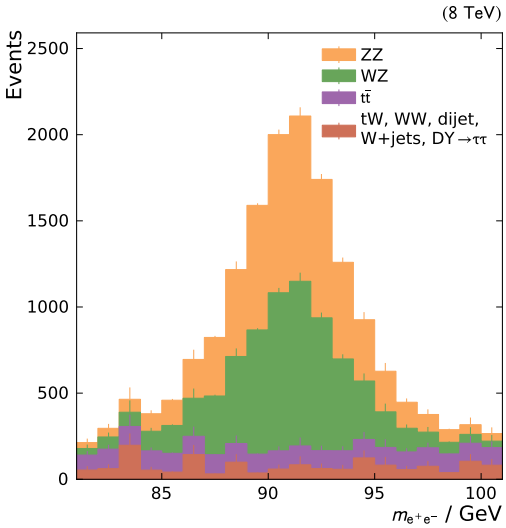


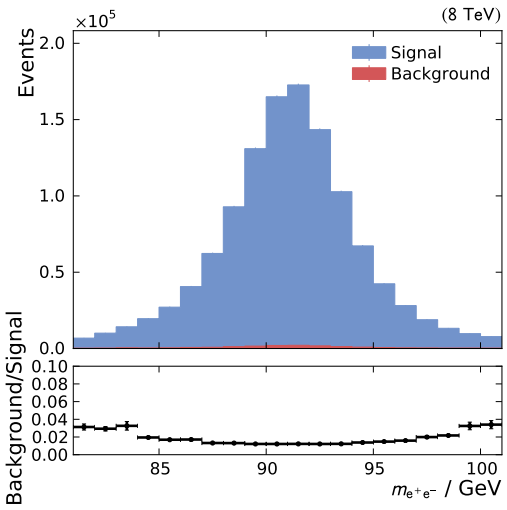
Figure 6.6: Exemplary Feynman diagrams for the most important background processes to Z(→ e⁺e⁻)+jet production.

Table 6.1: Cross sections of the studied background processes, used to normalize the samples of simulated events [18]. The program and perturbative order for each calculation is also shown. Further information on CMS-internal names and the sizes of the simulated samples is given in Appendix A.1.

Process	Program	Order	Cross section / pb
ZZ	MCFM 6.6 [132]	NLO	17.7
WZ	MCFM 6.6	NLO	33.2
WW	MCFM 6.6	NLO	54.8
$t\bar{t}$($\ell\ell$ decay channel)	Top++ 2.0 [133]	NNLO	26.8
tW	[134]	NNLO	22.2
$Z \to \tau^+\tau^-$ ($m_{\ell\ell} > 20\,\text{GeV}$)	FEWZ 3.1 [94]	NNLO	1966.7
W+jets	FEWZ 3.1	NNLO	37509.8
dijet ($20 < p_T < 30\,\text{GeV}$)	Pythia 6 [123]	LO	2.9×10^8
dijet ($30 < p_T < 80\,\text{GeV}$)	Pythia 6	LO	7.5×10^7
dijet ($80 < p_T < 170\,\text{GeV}$)	Pythia 6	LO	1.2×10^6
dijet ($170 < p_T < 250\,\text{GeV}$)	Pythia 6	LO	31274.0
dijet ($250 < p_T < 350\,\text{GeV}$)	Pythia 6	LO	4227.7
dijet ($p_T > 350\,\text{GeV}$)	Pythia 6	LO	803.5



(a) Invariant mass distribution, background processes only



(b) Ratio of expected background and signal contributions

Figure 6.7: Background contributions. Left: The largest contributions are WZ and ZZ processes. Right: The relative background slightly increases at the flanks of the Z boson mass peak. Overall, the relative background amounts to less than 2%.

6.2.5 Unfolding

Effects such as finite detector resolution or reconstruction imperfections bias the electron measurement and subsequently smear the distributions of Z boson p_{T} and rapidity. These effects have to be corrected for in a procedure known as *unfolding*.

A common approach to this problem is to formulate the effects of the reconstruction on a distribution (*folding*) as the result of a reversible operation. After detector effects are studied and determined, the folding can be reversed (unfolding). Thereby, it is possible to obtain the original or "true" distribution of a measured quantity in the absence of systematic biases.

The simplest approach is to relate the event count in each bin of the measured distribution m_i to the corresponding bin of the true distribution t_i via a single multiplicative factor c_i:

$$m_i = c_i t_i \,. \tag{6.8}$$

This is called the *bin-by-bin method*. The multiplicative approach of this procedure however neglects the additive nature of bin migrations: The migrations into one bin of the measured distribution originate from different bins of the true distribution. As a consequence, more sophisticated methods are necessary.

The measured and true distributions can be represented by the vectors $\vec{m}$ and $\vec{t}$, respectively. They are related to each other via the response matrix $\mathbf{R}$:

$$\vec{m} = \mathbf{R}\vec{t}. \tag{6.9}$$

$\mathbf{R}_{ij}$ denotes the fraction of events in t_j that "end up" in m_i as a result of the reconstruction. A perfect measurement would correspond to $\mathbf{R}_{ii} = 1$ and $\mathbf{R}_{ij} = 0$ for $i \neq j$, i.e. the response matrix is diagonal and there are no bin migrations.

The inverted response matrix is used to obtain the true distribution from the measured one. This is called the *matrix inversion method*. However, this can lead to biases in the case of large bin migrations or large statistical fluctuations. To deal with these fluctuations, the *iterative d'Agostini unfolding* [135] utilizes a regularization method which however leads to an additional bias.

Unfolding of Z(→ e⁺e⁻)+jet Cross Section

The response matrix entries are determined from simulation. A sample of simulated Z+jet events including a simulation of the CMS detector is used. The matrix element calculation was performed with MadGraph [93] while Pythia 6 Tune Z2* was used for the parton shower and Geant4 [95] for the detector simulation. Additional information on the simulated sample is given in Table A.1.

The same phase space selection is applied to both simulated and reconstructed particles to assure equal kinematic acceptance on both levels. The response matrix is constructed by filling the simulated events in bins according to the value of Z boson p_{T} or rapidity on particle-level and reconstructed level, i.e. before and after the detector simulation. After filling, the columns of the response matrix are normalized to one. The binning of the

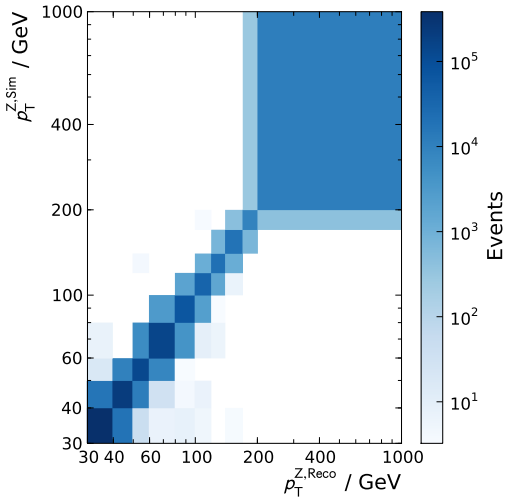
response matrix is extended compared to the measured distribution to correctly simulate the migration of events from outside the analysis phase space. With the obtained response matrix, the "true" distribution in data can be estimated from the measured one. The unfolding is performed with the RooUnfold package [136].

The response matrix for the unfolding of the Z boson p_{T} distribution is shown in Figure 6.8(a). The off-diagonal entries are sparsely populated, causing only small bin migrations in the unfolding procedure. Evidently, the excellent electron reconstruction and calibration leads to only minor smearing of electron p_{T}. A comparison of the distribution before and after unfolding is shown in Figure 6.8(b). The differences are small, at the level of only few percent.

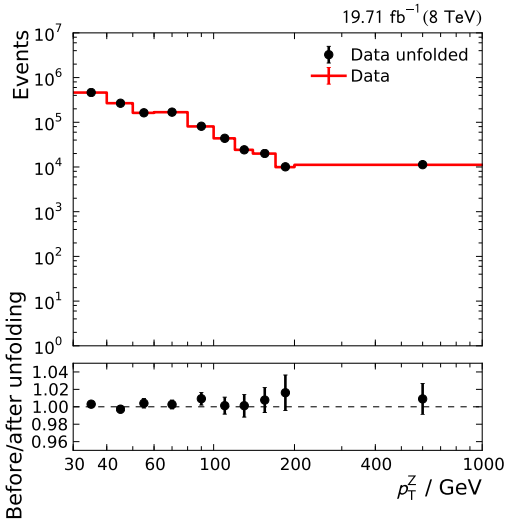
The bin-by-bin, matrix inversion and iterative d'Agostini methods are compared in Figure 6.8(c). The results for all three methods are in agreement. As both the (un)folding effects and the statistical uncertainty are small, the regularization approach of the iterative d'Agostini method does not lead to a significant improvement. For this reason, the more robust matrix inversion is chosen as the default unfolding method.

Another cross-check is performed with the response matrix constructed from a sample of simulated events where Powheg [137] was used for the matrix element calculation. The unfolding of the measured $p_{\mathrm{T}}^{\mathrm{Z}}$ distribution was performed for both response matrices and the resulting distributions are compared, see Figure 6.8(d). No significant difference was found.

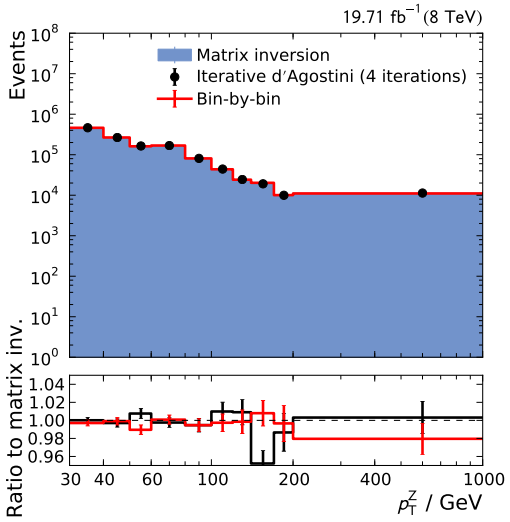
The corresponding plots for the unfolding of the $|y_{\mathrm{Z}}|$ distribution and the double-differential measurement are shown in Figure A.15 and Figures A.16 to A.19. The rapidity measurement is even more precise than the energy measurement which is limited by the ECAL resolution.



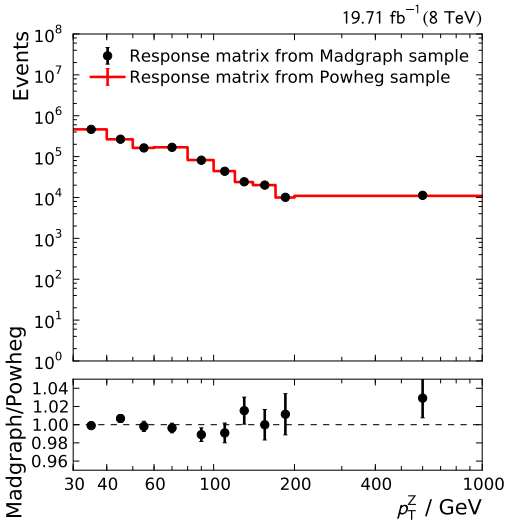
(a) Response matrix



(b) Comparison of distribution before and after unfolding



(c) Result of unfolding with different methods



(d) Result of unfolding with response matrices from different samples

Figure 6.8: Response matrix used for the unfolding of the p_T^Z distribution (upper left). p_T^Z distribution before and after unfolding (upper right). Comparison of different unfolding methods (lower left). Result of the unfolding with the response matrix constructed from different simulated samples (lower right).

6.3 Determination of Systematic Uncertainties

In this section, the evaluation of systematic uncertainties, both experimental and theoretical, is described.

6.3.1 Experimental Uncertainties

The following systematic uncertainties were considered:

Luminosity The uncertainty on the integrated luminosity measured by the CMS detector is 2.6% for the data taken at $\sqrt{s} = 8\,\text{TeV}$ [73]. It is constant for all bins.

Statistical This uncertainty comprises the statistical uncertainty of the data and the response matrix. It is estimated via a series of 10 000 pseudo-experiments ("Toy Monte Carlos"). For each pseudo-experiments, the entries of the data histogram and the response matrix are independently but simultaneously varied. The variations are random according to a Gaussian distribution with mean and standard deviation corresponding to the number of bin entries and square root of the number of bin entries.
Thus, the unfolding yields a slightly different result for each pseudo-experiments. The central result and the associated statistical uncertainty are then determined via averaging over all pseudo-experiments and calculating the mean and standard deviation for each bin. A cross-check with different unfolding methods or the response matrix constructed from different simulated samples (see Figure 6.8(b) bottom row) gave no indication for an additional effect. No systematic uncertainty was attributed. The statistical correlations between the bins were also determined (see Figure A.19) and taken into account for PDF fits. The resulting statistical uncertainty results mainly from the uncertainty in the data and the response matrix statistical uncertainty contributes only marginally, see Figure A.20.

Reconstruction, identification and trigger efficiency The respective scale factors are varied up- and downwards according to the respective statistical and systematic uncertainty, see [131].

Electron energy scale The p_T of each electron is varied up- and downwards according to the electron scale uncertainty depending on electron p_T and $|\eta|$, see [57].

Backgrounds The uncertainty on the cross section due to background processes is estimated by varying the cross sections of the background processes up- and downwards by 50% for background subtraction.

For the latter three uncertainty sources, the upwards/downwards variations are propagated to the final cross section. Each uncertainty is defined as the maximum difference of any of the two variations compared to the central result.

All uncertainties are treated as uncorrelated. The obtained uncertainties in each bin are added in quadrature to obtain the total experimental uncertainty. The results are

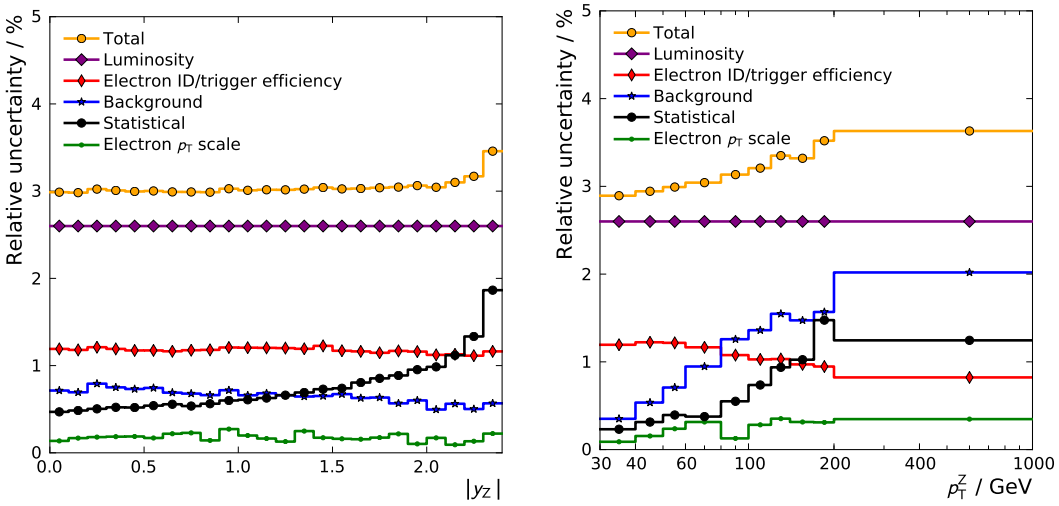


Figure 6.9: Experimental uncertainties as a function of Z boson rapidity (left) and p_T (right). The total uncertainty is obtained by quadratically adding the individual uncertainty sources. The luminosity uncertainty is the dominant uncertainty. Statistical uncertainties increase at higher rapidity and higher p_T, where also the background uncertainty increases. The total uncertainty is at the level of 3% to 4%.

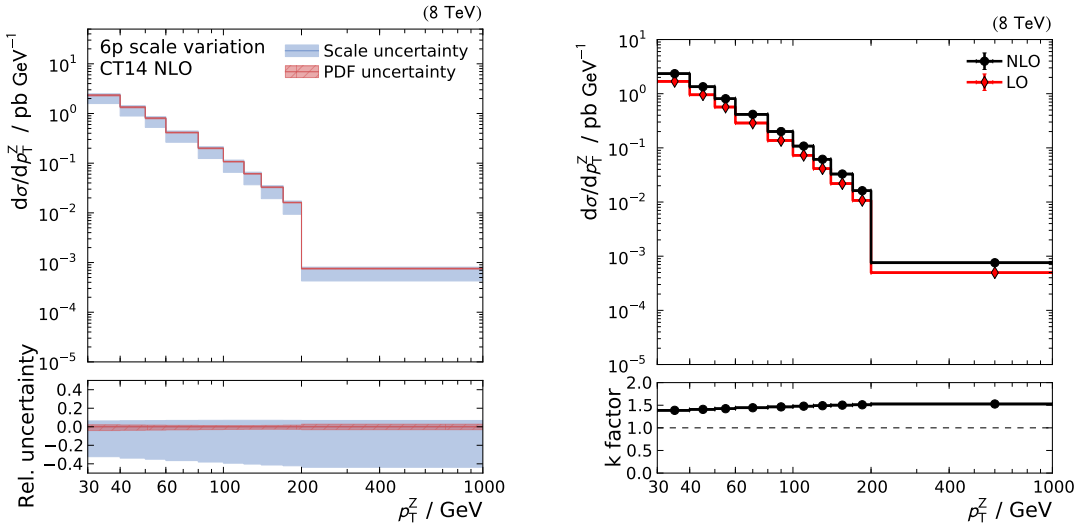
shown in Figure 6.9 for the rapidity and p_T distributions of the Z boson. The uncertainty of the luminosity measurement is the dominant contribution to the total uncertainty. The uncertainties for the efficiency scale factors are slightly above 1% with a decrease towards higher p_T. For higher Z boson p_T, the uncertainty from the background modelling increases from 0.4% to 2% because of the higher relative contribution from high-energetic diboson and $t\bar{t}$ events. In total, the experimental uncertainty is at the level of only 3% to 4%.

For the double-differential measurement, the experimental uncertainties are shown in Figure A.23. Other sources of uncertainty such as final-state radiation or Z boson polarization have been investigated in [138] and were found to be negligible compared to the leading uncertainties.

6.3.2 Theory Uncertainties

For the theory prediction, the scale and PDF uncertainties were estimated. In addition, the predictions for LO and NLO accuracy were compared.

The scale uncertainty has been determined by varying the factorization and renormalization scales μ_f and μ_r (see Section 2.1) independently by a factor 2 up or down, i.e. considering the multiplicative values 0.5, 1 and 2. The extreme combinations for (μ_f, μ_r) of (0.5, 2) and (2, 0.5) are omitted. The resulting six variations are (0.5, 0.5), (0.5, 1), (1, 0.5), (1, 2), (2, 1) and (2, 2). The largest difference between any variation and the central value determines the scale uncertainty.



(a) Scale uncertainty of the NLO cross section prediction

(b) Comparison of cross section predictions at LO and NLO accuracy

Figure 6.10: Left: Theory prediction with scale and PDF uncertainties. Right: Theory predictions at leading- and next-to-leading-order accuracy. The ratio between both predictions, the so-called *k factor*, is also shown.

The result is shown in Figure 6.10(a) for the p_T distribution. For the rapidity distribution or the double-differential measurement, the results are shown in Figure A.21. The scale uncertainty is highly asymmetric and increases with p_T. In the highest p_T bin, the downwards scale uncertainty amounts to over 40%.

To further study the dependence on the truncation of the perturbative series, the cross section predictions at LO and NLO accuracy are compared. The ratio between the cross section calculated at LO and NLO accuracy is commonly called the *k factor*:

$$k = \frac{\sigma_{\mathrm{NLO}}}{\sigma_{\mathrm{LO}}} \,. \tag{6.10}$$

The k factors in each bin of rapidity and p_T are shown in Figure 6.10(b). As expected, they are larger where the scale uncertainty is also larger, i.e. at higher p_T. The k factors are around 1.4 to 1.5, confirming the necessity of a calculation at (at least) NLO accuracy.

Recent theory results [139] suggest that the cross section differences between NLO and NNLO are at the level of only few percent. This means that the effect of higher orders in the cross section calculation is rather small. The reliability of the theory prediction would certainly profit from the increased accuracy at NNLO.

The uncertainty on the cross section introduced by the uncertainty of the PDFs is shown in Figure 6.10(a) for the CT14 [13] PDF set. Ranging from 3% to 4%, it is at a comparable level with total experimental uncertainty.

6.4 Comparison of Measurement and Simulation

After efficiency corrections, background subtraction and unfolding, the results of the measurement can be compared to the theory predictions. A comparison of data with predictions from different PDF sets can be seen in Figure 6.11 for the single-differential measurements in rapidity (left) and p_{T} (right). All PDF sets describe the data over most of the studied phase space. The ABM12 prediction is few percent higher than the other predictions in the low rapidity and low p_{T} region. Differences can be found at higher rapidities, especially at $|y_{\mathrm{Z}}| > 2.0$, or at highest p_{T}. However, these phase space regions are only sparsely populated.

For the double-differential measurement in p_{T} and $|y|$, the results are shown in Figure 6.12. As was already seen for the single-differential measurement, the simulation reliably describes the data in most studied phase space regions. Increasing differences can be found in the outer rapidity bin and in the bin of highest p_{T}. Comparisons of data with different PDF sets for the double-differential measurement are shown in Figure A.22.

In a measurement of the Z boson p_{T} distribution at $\sqrt{s} = 7\,\mathrm{TeV}$ with the ATLAS detector [140], a similar disagreement for the SHERPA prediction was found. The disagreement at higher p_{T} possibly originates from the lack of electroweak corrections in the cross section calculation. A recent study [141] shows that electroweak corrections decrease the cross section at higher p_{T} which could explain the observation. Recent findings [142] suggest that non-perturbative effects could also affect the Z boson p_{T} in Z+jet events. The deviations at higher rapidity are possibly caused by imperfections in the calculation, e.g. from inaccurate PDFs. As the reconstruction of electrons is more difficult in the endcap section of the detector, this phase space region is also challenging from an experimental point of view.

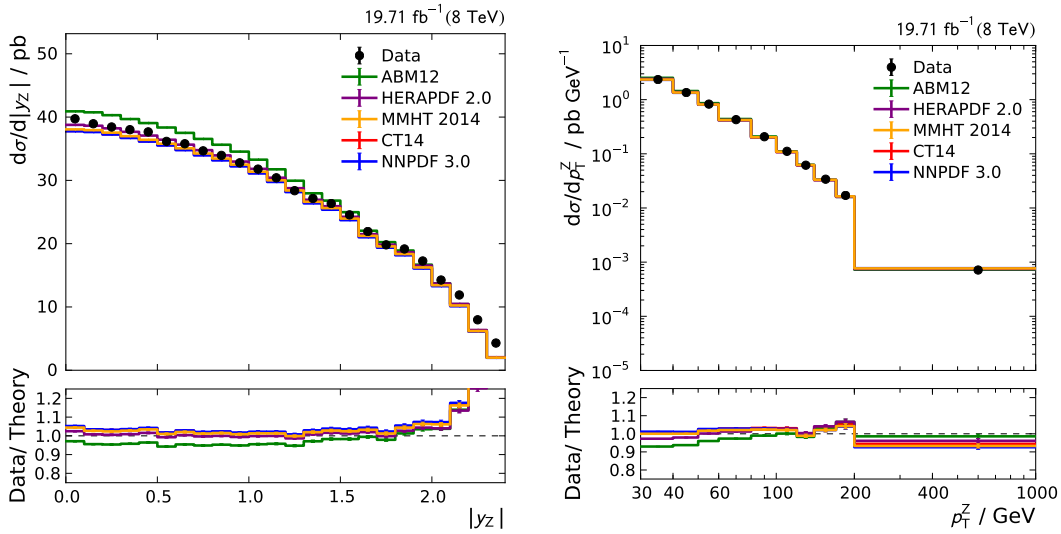


Figure 6.11: Comparison of data and theory predictions calculated with different PDF sets. The theory well describes the data for most of the studied phase space. Increasing differences can be seen for higher absolute rapidity or higher p_{T}. The prediction from ABM12 differs a few percent compared to the other PDF sets.

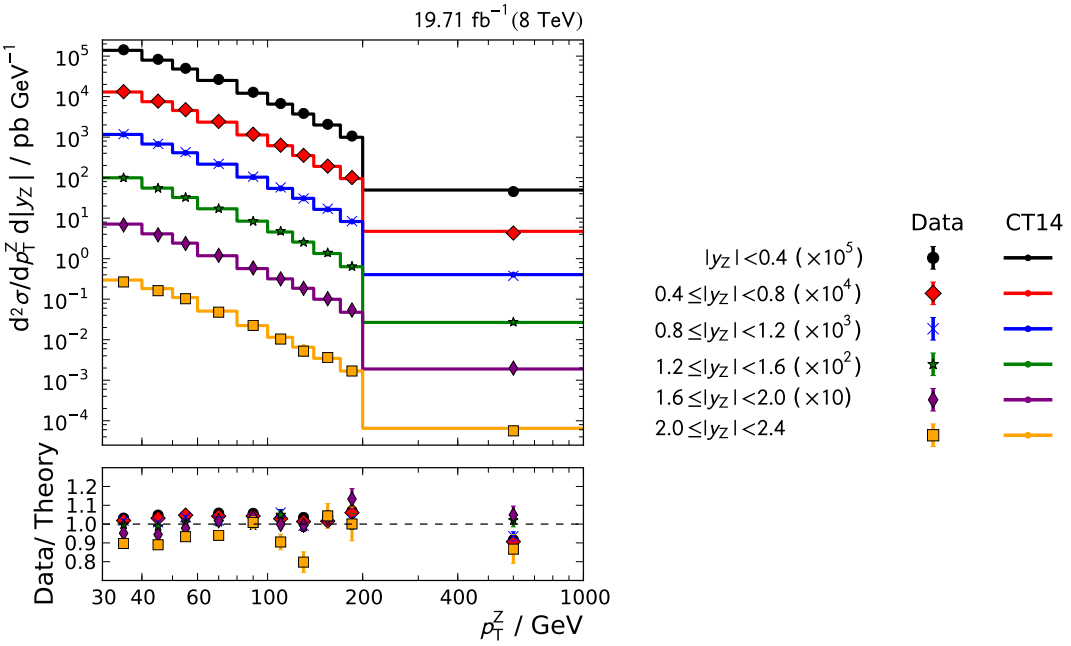


Figure 6.12: Z boson double-differential cross section in p_{T} and absolute rapidity. The simulation well describes the data over most of the studied phase space. Differences between data and simulation arise at higher rapidity and higher transverse momentum.

6.5 PDF Determination with xFitter

The results of the double-differential measurement as seen in Figure 6.12 are used to determine the PDFs. All statistical and systematic uncertainties including their correlations are taken into account. To consider only phase space regions which are reliably described, the bins of highest p_{T} and highest rapidity are not included in the fit. The excluded phase space regions correspond to less than 4.6% of the events of the double-differential measurement.

The combined HERA I+II data of charged and neutral current $e^{\pm}p$ deep-inelastic scattering at different beam energies published in [9] are also used[2]. Feynman diagrams for these processes are shown in Figure 2.3. The data correspond to an integrated luminosity of $1\,\mathrm{fb}^{-1}$. These data are the basis for the HERAPDF 2.0 set of parton distribution functions.

By comparing data and theory, the PDFs can be constrained. The x-dependence of the PDFs is parametrized and the cross section is calculated based on the PDFs. The PDF parameters are then fitted to minimize the χ^2 between theory and HERA in combination with CMS $\mathrm{Z}(\rightarrow \mathrm{e}^+\mathrm{e}^-)$+jet data. The result is compared with a fitting procedure using only HERA data. Both the change in the PDFs and in their uncertainties is studied.

The fitting of the PDF parameters is performed in accordance with [9]. The analysis strategy is similar to studies using CMS jet data [4, 143]. xFitter, as described in Section 4.3, is used to perform the PDF determination.

6.5.1 Combined PDF Fit with HERA and CMS Data

In accordance with HERAPDF 2.0 [9], the PDF parametrization at the starting scale Q_0^2 is

$$xg(x) = A_g x^{B_g}(1-x)^{C_g} - A'_g x^{B'_g}(1-x)^{C'_g}\,, \tag{6.11}$$

$$xu_v(x) = A_{u_v} x^{B_{u_v}}(1-x)^{C_{u_v}}(1+E_{u_v}x^2)\,, \tag{6.12}$$

$$xd_v(x) = A_{d_v} x^{B_{d_v}}(1-x)^{C_{d_v}}\,, \tag{6.13}$$

$$x\bar{U}(x) = A_{\bar{U}} x^{B_{\bar{U}}}(1-x)^{C_{\bar{U}}}(1+D_{\bar{U}}x)\,, \tag{6.14}$$

$$x\bar{D}(x) = A_{\bar{D}} x^{B_{\bar{D}}}(1-x)^{C_{\bar{D}}} \tag{6.15}$$

with the gluon PDF xg and the valence quark PDFs xu_v and xd_v. $x\bar{U}$ and $x\bar{D}$ denote the up-type and down-type antiquark distributions, i.e. $x\bar{U} = x\bar{u}$ and $x\bar{D} = x\bar{d} + x\bar{s}$.

The normalization parameters A are constrained by quark number and momentum sum rules. $B_{\bar{U}}$ is set equal to $B_{\bar{D}}$ to enfore a consistent low-x behaviour for sea quarks. The strange quark PDF is assumed to be proportional to the d antiquark PDF, i.e. $x\bar{s} = f_s x\bar{D}$ at Q^2_{min} with the fixed parameter $f_s = 0.4$. To further ensure reasonable behaviour of the PDFs, the relation $A_{\bar{U}} = A_{\bar{D}}(1-f_s)$ is enforced. C_g is fixed to 25 such that the negative gluon term only contributes at small x. This leaves 15 free parameters to be fitted.

[2] These data are referred to as 'HERA data' below.

The parameters of the physics model, e.g. the strange quark fraction, the strong coupling constant or the heavy quark masses, are set in accordance with [9]. The heavy flavours are treated in the general-mass variable-flavour-number scheme [144]. The PDFs are parametrized at the starting scale $Q_0^2 = 1.9\,\mathrm{GeV}^2$ and evolved to the scale of the measurement via the DGLAP equations (see Section 2.1.2) at NLO, the same perturbative order as used for the Z(→ e⁺e⁻)+jet cross section calculation. xFitter relies on QCDNUM [105] for the evolution of the PDFs and calculation of DIS cross sections. For the CMS Z(→ e⁺e⁻)+jet data, the cross sections are calculated by convolving the PDFs with the hard scattering coefficients stored in the fastNLO tables.

6.5.2 Definition of Goodness-of-Fit Estimator

The parameters of the PDFs are fitted to minimize the goodness-of-fit estimator χ^2. This step is performed with xFitter which relies on the Minuit [106] minimization tool. All statistical uncertainties as well as correlated and uncorrelated systematic uncertainties have to be properly taken into account. For a data set D consisting of the measurements d_i and the corresponding theory predictions t_i the definition is

$$\chi^2_D = \chi^2_{D,\mathrm{data}} + \sum_\alpha n_\alpha^2 \tag{6.16}$$

$$= \sum_{ij} \left(t_i - \sum_\alpha \Gamma^i_\alpha(t_i) n_\alpha - d_i \right) \mathbf{C}^{-1}_{\mathrm{stat},ij}(d_i, d_j) \left(t_j - \sum_\alpha \Gamma^j_\alpha(t_j) n_\alpha - d_j \right) + \sum_\alpha n_\alpha^2 \,. \tag{6.17}$$

The sum runs over all data points i, j. The simple difference between theory and data $(t_i - d_i)$ is extended by a term $\Gamma^i_\alpha(t_i) n_\alpha$ accounting for systematic uncertainties. Γ_α is the systematic uncertainty arising from an uncertainty source α. The nuisance parameters n_α quantify the shift of the prediction caused by each systematic uncertainty.

The shifts of the systematic uncertainties are applied to the theory prediction to avoid the bias from multiplicative uncertainties applied to data. The covariance matrix $\mathbf{C}_{\mathrm{stat}}$ takes into account the statistical correlation introduced by the unfolding. However, the correlations for off-diagonal entries are marginal (see Figure A.19). $\sum_\alpha n_\alpha^2$ is a "penalty" term: Since the nuisance parameters shift the prediction in terms of standard deviations, they have to be added to the χ^2 quadratically.

The total χ^2 for N data sets includes the penalty term only once and is defined as

$$\chi^2_{\mathrm{tot}} = \sum_D^N \chi^2_{D,\mathrm{data}} + \sum_\alpha n_\alpha^2 \,. \tag{6.18}$$

6.5.3 Determination of PDF Uncertainties

Three PDF uncertainty sources have to be considered:

Experimental uncertainties The statistical and systematic uncertainties of the cross section measurement are propagated to the PDFs via the *Hessian method* [145]:

In the parameter space spanned by the n PDF parameters, a one-σ confidence level on the PDFs corresponds to $\Delta\chi^2 = 1$ around the best-fit minimum χ^2_{min}. As the n PDF parameters are correlated, the parameter basis is transformed into an orthonormal eigenvector basis. The eigenvectors are independently from each other varied positively and negatively until $\chi^2 - \chi^2_{\text{min}} = 1$. The result of each variation is a point in the eigenvector space which corresponds to a PDF set different from the central result. From all variations, a set of $2n$ eigenvector-variation PDFs is obtained. The experimental uncertainty on the PDFs is then determined by quadratically adding the upwards or downwards differences between the predictions from the central PDF and the variations.

Model uncertainties This uncertainty arises from the uncertainty of certain parameters of the physics model used in the fitting procedure. To estimate their impact, fits with variations of the model parameters were performed and compared to the central fit. The masses of b and c quarks M_b and M_c, the strange quark fraction f_s, the strong coupling constant $\alpha_\text{s}(M_\text{Z})$ and the minimum Q^2 of HERA data included in the fitting procedure, Q^2_{min}, is varied. The central values of the parameters and the upwards- and downwards-variations are given in Table 6.2.

Parametrization uncertainties A more general parametrization of the PDFs with additional parameters D and E is used, fitted and compared to the default parametrization, see Equations 6.11–6.15. The extended parametrization is

$$xg(x) = A_g x^{B_g}(1-x)^{C_g}(1 + D_g x + E_g x^2) - A'_g x^{B'_g}(1-x)^{C'_g}, \tag{6.19}$$

$$xf(x) = A_f x^{B_f}(1-x)^{C_f}(1 + D_f x + E_f x^2) \tag{6.20}$$

where $xg(x)$ is the gluon PDF and $xf(x)$ indicates the parametrization for all quark flavours. The parametrization extension and fitting is performed for each PDF and parameter independently. The maximum difference to the default parametrization is defined as parametrization uncertainty. In addition, the starting scale Q_0 of the fit is varied.

The individual uncertainties are added in quadrature to obtain the total PDF uncertainty.

6.5.4 Results of PDF Parameter Fitting

For the combined fit to HERA and CMS $\text{Z}(\rightarrow \text{e}^+\text{e}^-)$+jet data, the values for the fitted parameters are shown in Table 6.5. The χ^2 values of the individual data sets after the fitting procedure are shown in Table 6.3 along with the result of fitting to only HERA data. The total χ^2 values are reported in Table 6.4.

For $|y_\text{Z}| < 1.6$, the χ^2 per data point is close to one, indicating compatibility of the theory and the data. Increasing differences between data and simulation at rapidities beyond 2.0 were already described above. For $1.6 < |y_\text{Z}| < 2.0$, the large χ^2 indicates a clear deviation whose origin may be related to the areas of PDFs, experimental systematics or theory calculation. No conclusive answer was found and future studies are required.

Table 6.2: Parameter values and variations to determine model and parametrization (Q_0) uncertainties. The central value and variations follow the procedure in [9].

	Parameter values		
Parameter	Central	Down	Up
f_s	0.4	0.3	0.5
$\alpha_s(M_Z)$	0.118	0.117	0.119
M_c / GeV	1.47	1.41	1.53
M_b / GeV	4.5	4.25	4.75
Q^2_{min}/ GeV^2	3.5	2.5	7.5
Q^2_0 / GeV^2	1.9	1.6	2.2

Table 6.3: χ^2 between data and theory for the individual data sets after PDF determination.

Data		n_{data}	HERA χ^2_{data}/n_{data}	HERA+CMS χ^2_{data}/n_{data}
HERA I+II NC e⁺p	$E_p = 920\,\text{GeV}$	377	1.104	1.149
	$E_p = 820\,\text{GeV}$	70	0.934	0.965
	$E_p = 575\,\text{GeV}$	254	0.834	0.852
	$E_p = 460\,\text{GeV}$	204	1.029	1.049
HERA I+II CC e⁻p		42	1.183	1.169
HERA I+II NC e⁻p		159	1.389	1.399
HERA I+II CC e⁺p		39	1.087	1.367
CMS Z(→ e⁺e⁻)+jet	$\lvert y_Z\rvert < 0.4$	9		1.320
	$0.4 \leq \lvert y_Z\rvert < 0.8$	9		1.161
	$0.8 \leq \lvert y_Z\rvert < 1.2$	9		1.393
	$1.2 \leq \lvert y_Z\rvert < 1.6$	9		1.273
	$1.6 \leq \lvert y_Z\rvert < 2.0$	9		5.193

Table 6.4: Total χ^2 between data and theory after PDF determination.

Data	n_{dof}	χ^2_{tot}/n_{dof}
HERA I+II	1131	1.174
HERA I+II and CMS Z(→ e⁺e⁻)+jet	1176	1.255

Table 6.5: Resulting PDF parameter values from the combined fit of HERA and CMS Z($\rightarrow$ e^+e^-)+jet data.

	A	B	C	D	E	A'	B'	C'
xg	4.60	-0.03	8.97			1.76	-0.12	25.00
xu_v	4.21	0.72	4.72		11.74			
xd_v	3.45	0.83	4.21					
$x\bar{U}$	0.11	-0.17	6.59	9.83				
$x\bar{D}$	0.19	-0.17	11.06					

The resulting PDFs and their experimental, model and parametrization uncertainties are shown in Figure 6.13 and Figure 6.14 for $Q^2 = 1.9\,\mathrm{GeV}^2$. The results from the fit of only HERA data (left column) and the combined HERA and CMS Z($\rightarrow$ e^+e^-)+jet data (right column) are presented. For all studied flavours, the experimental uncertainties decrease as is expected with the addition of more constraining data. The model and parametrization uncertainties also decrease for most flavours. The model uncertainties also contain the variation in $\alpha_s(M_Z)$ which however lead to only small PDF variations. For the gluon PDF, the decrease of total uncertainty at medium-x to low-x is mainly caused by the decrease of the parametrization uncertainty. For the u valence, d valence and sea quark PDFs the parametrization uncertainty shows some "bumps" at medium- or high-x which decrease with the inclusion of CMS Z($\rightarrow$ e^+e^-)+jet data.

A direct comparison of the results from the combined fit and the HERA-only fit is shown in Figure 6.15. The total uncertainty decreases for all flavours over the entire x-range. For the gluon PDF, the reduction of uncertainty is especially prominent. A significant PDF change is observed for the gluon and in the medium-to-high x region for the sea quark PDF. The gluon PDF changes towards a "harder" gluon, i.e. the PDF decreases at medium x and increases at high x. For the gluon PDF, the change towards a harder gluon is compatible with similar studies analysing CMS jet data [4].

Using a similar approach, the impact of Z boson measurements with Tevatron data on the PDFs is described in [146]. In contrast, the CMS data provides a much better handle on the gluon distribution due to the large contribution of the qg production channel, see Figure 6.1(b). In addition, the higher centre-of-mass energy at the LHC increases the range in Q. Visualizations of the fitted PDFs for other flavours or higher Q^2 can be found in Figures A.24 to A.26.

A comparison of the PDFs determined from HERA data, from the combined HERA and CMS Z($\rightarrow$ e^+e^-)+jet data, and PDFs from the CTEQ and NNPDF collaborations is shown in Figure 6.16. The inclusion of CMS data decreases the difference between the PDFs from HERA and the CT14 and NNPDF 3.0 sets, especially for the gluon. For the CTEQ and NNPDF collaborations, the PDF determination is based on data from deep-inelastic scattering but also includes measurements by CMS, ATLAS and LHCb at $\sqrt{s} = 7\,\mathrm{TeV}$. Therefore, the observed shift of the PDFs with the inclusion of CMS Z($\rightarrow$ e^+e^-)+jet data in the own fit agrees with the general trend induced by LHC data.

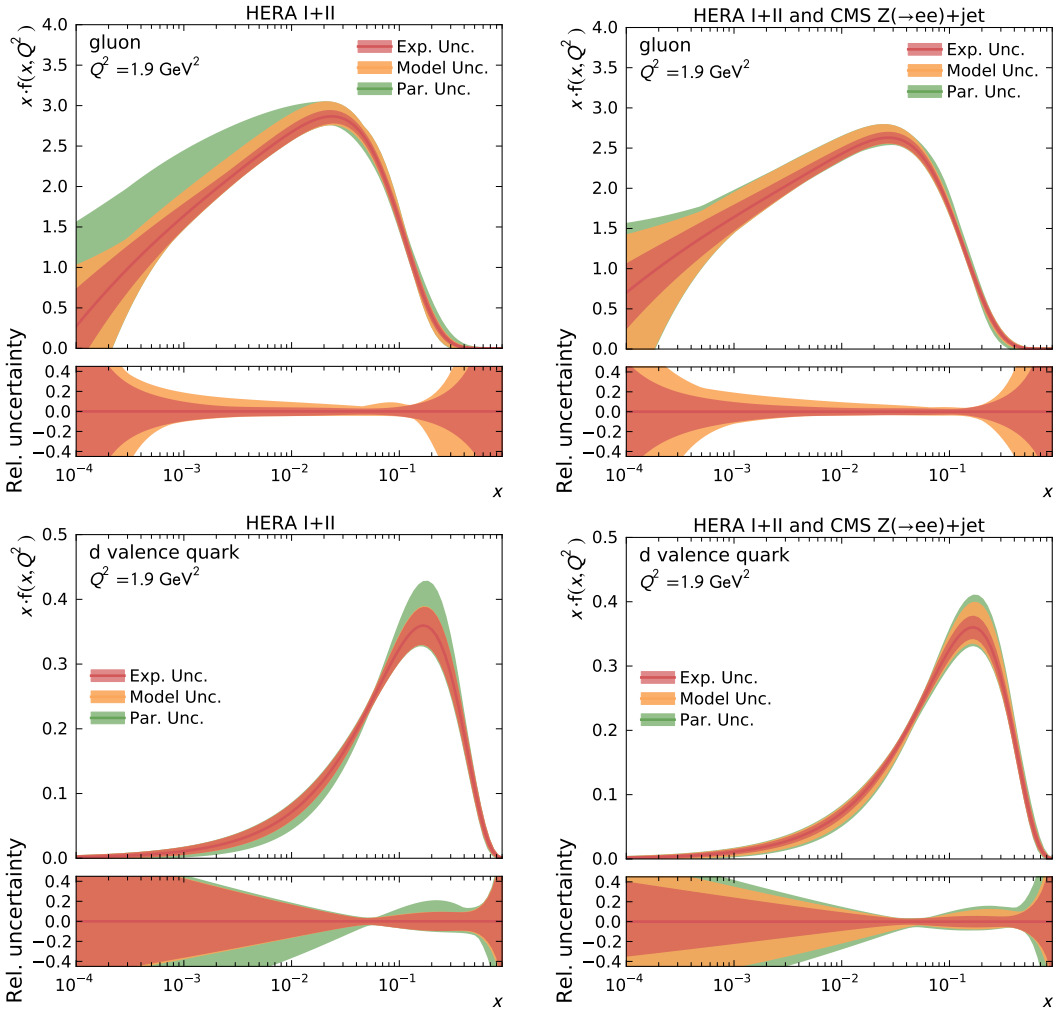


Figure 6.13: PDF uncertainties. The PDFs determined from HERA (left column) and the combined HERA and CMS data (right column) are compared. The PDFs for gluon (top row) and d valence quark (bottom row) are shown. Experimental, model and parametrization uncertainties are added in quadrature. The inclusion of Z(→ e⁺e⁻)+jet data leads to a decrease in PDF uncertainties.

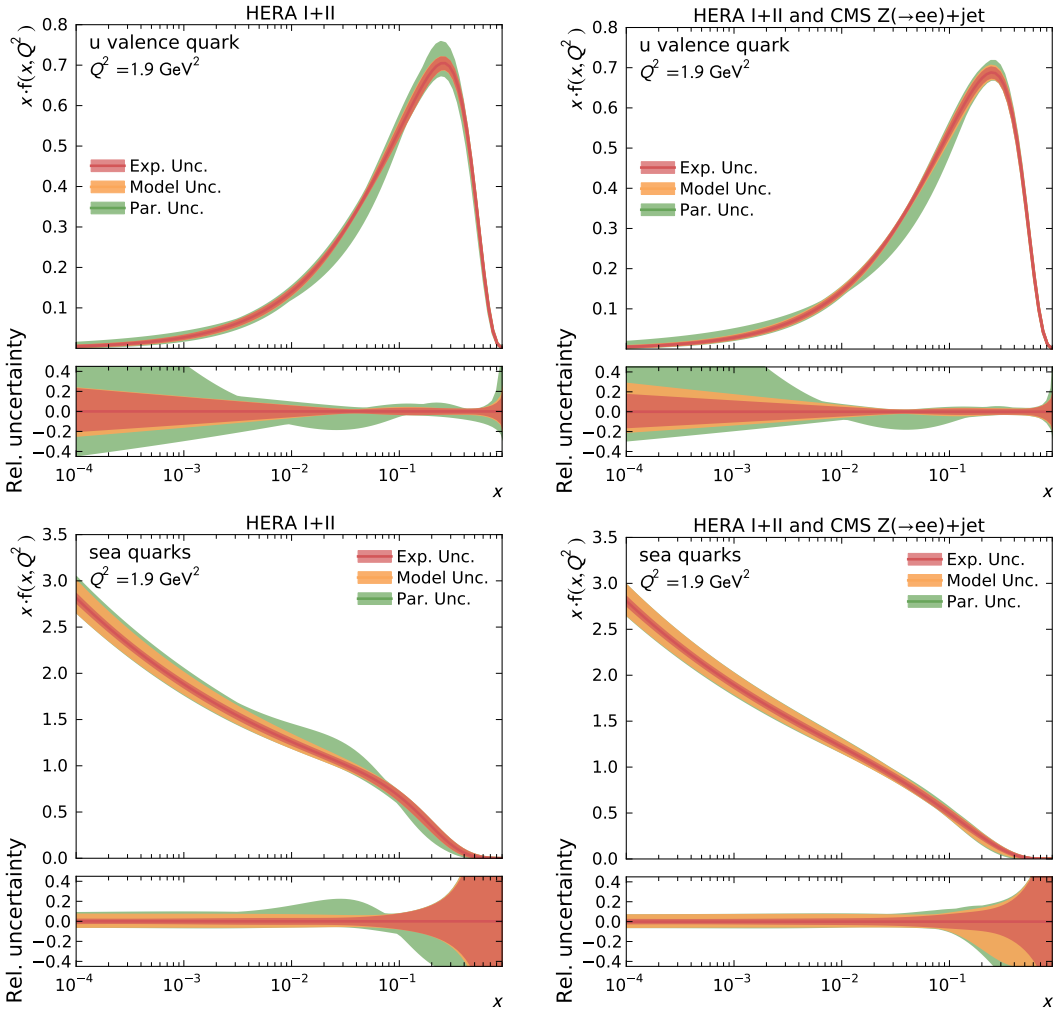


Figure 6.14: PDF uncertainties. The PDFs determined from HERA (left column) and the combined HERA and CMS data (right column) are compared. The PDFs for u valence quark (top row) and sea quarks (bottom row) are shown. Experimental, model and parametrization uncertainties are added in quadrature. The inclusion of Z($\rightarrow e^+e^-$)+jet data leads to a decrease in PDF uncertainties.

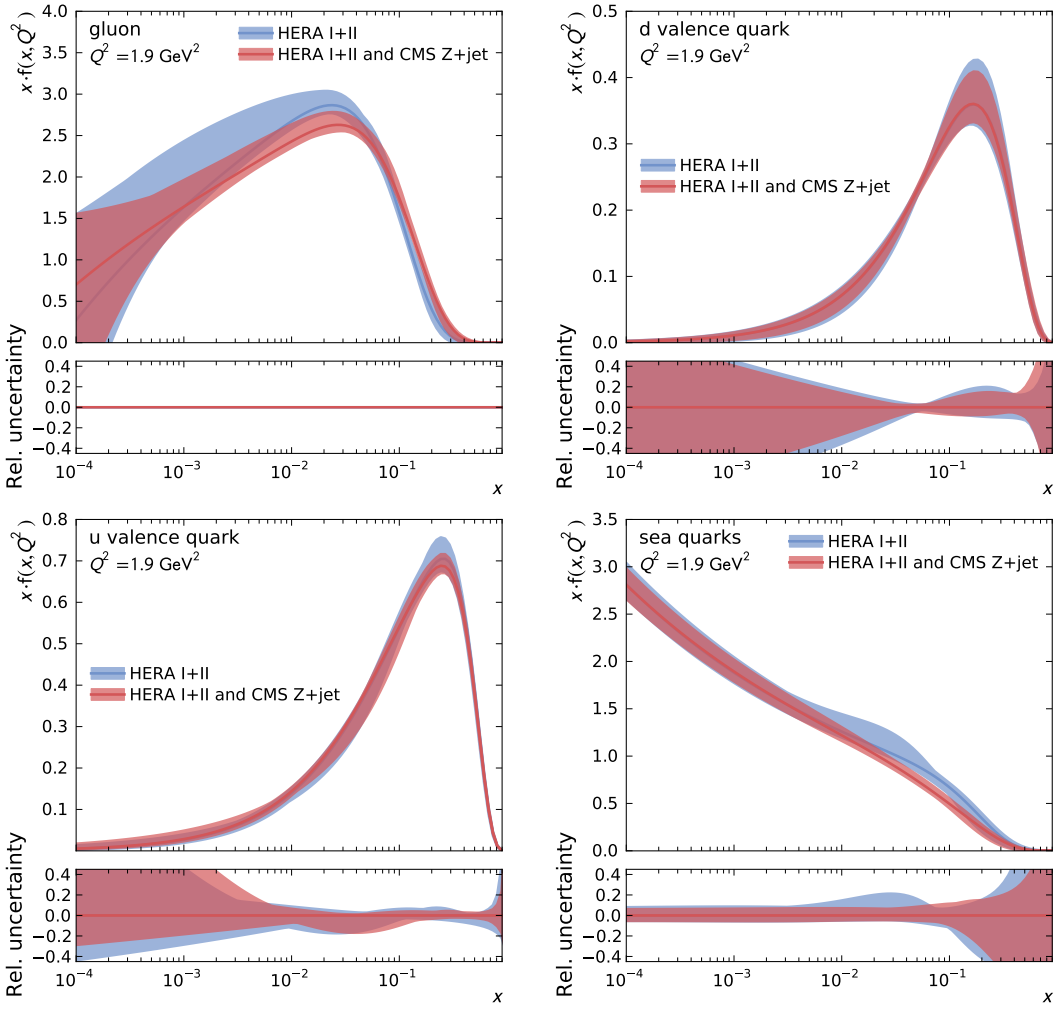


Figure 6.15: Comparison of PDFs with total uncertainties determined from only HERA data and the combination of HERA and CMS Z(→ e⁺e⁻)+jet data. With the inclusion of Z(→ e⁺e⁻)+jet data in the fit, the uncertainties are significantly reduced. A change of the PDF shape is observed for gluon and sea quarks.

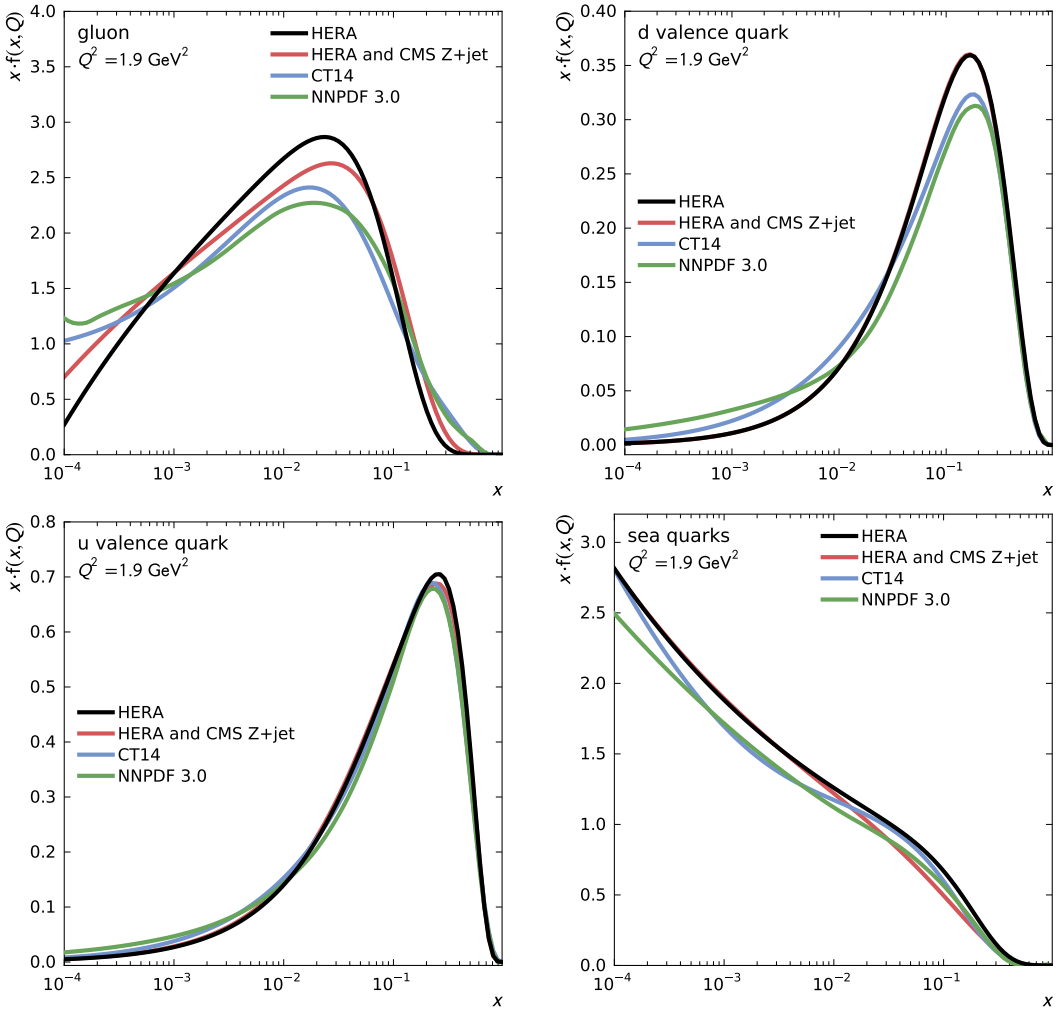


Figure 6.16: Comparison of PDFs determined from own fit and from PDF collaborations. With the inclusion of CMS Z($\rightarrow e^+e^-$)+jet data in the fit to HERA data, the PDFs shift closer to the PDFs from the CTEQ and NNPDF collaborations.

6.6 Summary

The Z(→ e⁺e⁻)+jet cross section was measured differentially in rapidity and transverse momentum of the reconstructed Z boson. A precise correction of the electron momentum was followed by efficiency corrections for electron reconstruction and identification. The expected contributions from background processes were estimated and the measured distributions were unfolded. The systematic uncertainties and their correlations were carefully determined. The total experimental uncertainty was found to be at the level of only few percent.

Theory predictions were calculated at next-to-leading-order accuracy based on different PDF sets. The contributions from the different production channels and the sensitivity of the cross section to the PDFs were determined. The scale and PDF uncertainties of the calculation were also studied.

The theory prediction agrees with the data for most of the studied phase space. Differences arise for the sparsely populated phase space regions of higher Z boson p_{T} and higher rapidity. These observations could be attributed to an imperfect calculation or the underestimation of systematic uncertainties and are similar to the findings of other Z boson p_{T} measurements performed by the ATLAS and CMS collaborations, see [138, 140].

The PDFs were determined with a combined fit to HERA and CMS Z(→ e⁺e⁻)+jet data. The uncertainties on the PDFs, especially those on the gluon PDF, were reduced compared to a fit using only HERA data. A significant change in the gluon PDF towards a harder gluon is observed, confirming the results of previous PDF studies by the CMS and ATLAS collaborations using jet data, see [4, 147]. A comparison of all determined PDFs is shown in Figure 6.17.

With the improved knowledge of the proton structure, the uncertainty in cross section calculations for proton collisions can be decreased. Even larger PDF constraints can be expected if the results from several analyses are used for the fit, e.g. a combination of the results from the electron and muon decay channels.

For Run 2 of the LHC, proton-proton collisions occur at $\sqrt{s} = 13\,\mathrm{TeV}$. This increases the Z+jet production rate and opens up new phase space regions. To better exploit the kinematic properties of the collision, the rapidity of the jet can also be taken into account. This allows to study both outgoing objects with high precision as the rapidity distribution is not affected by the comparatively large jet energy uncertainties. Additional progress in theoretical physics might also increase the accuracy of the theory calculation. A more precise luminosity measurement could further reduce the experimental uncertainties. As the Z+jet topology contains at least one QCD vertex, a determination of α_{s} is also possible.

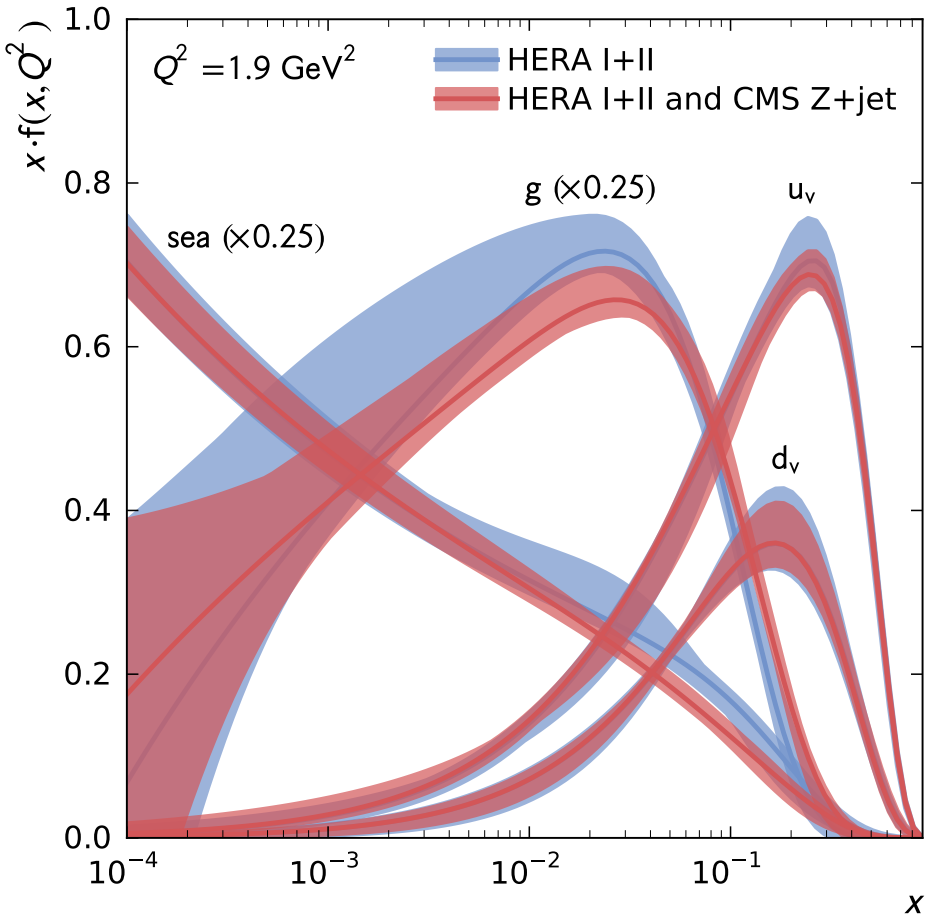


Figure 6.17: Comparison of PDFs determined from HERA and HERA+CMS data. The gluon and sea quark PDFs are scaled to be in comparable range with the valence quark PDFs. The inclusion of CMS Z($\rightarrow e^+e^-$)+jet data into the PDF fit shifts the gluon PDF and decreases PDF uncertainties.

7 Conclusion

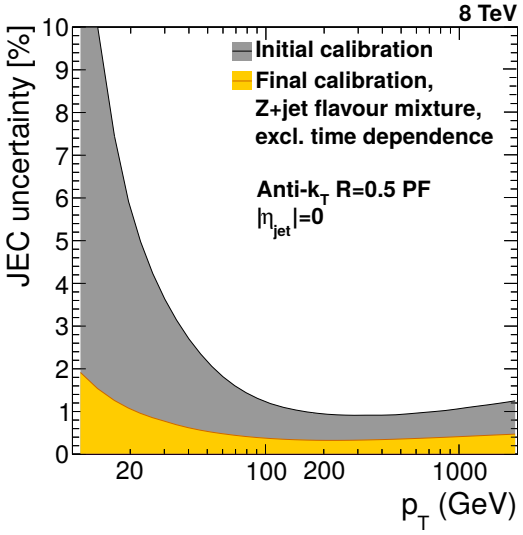
For the results presented in this thesis, the entire data recorded with the CMS detector at $\sqrt{s} = 8\,\text{TeV}$, corresponding to $19.7\,\text{fb}^{-1}$, was analysed. The high centre-of-mass energy and the large number of Z+jet events have enabled precision studies even in previously inaccessible phase space regions. The high performance in muon and electron reconstruction and the small background contributions have resulted in small experimental uncertainties. In this thesis, Z+jet events were utilized in studies of jet energy scale and proton structure.

For jet calibration, $\text{Z}(\rightarrow \mu^+\mu^-)$+jet events were used for a data-driven determination of the jet energy scale. Jet response and resolution were measured with two complementary methods. Numerous cross-checks were performed and state-of-the-art analysis techniques were explored to increase the precision and reliability of the calibration. Multivariate methods allowed a data-driven estimation of the differences between jets originating from different parton flavours. As a result, a thorough understanding of jet reconstruction has been obtained.

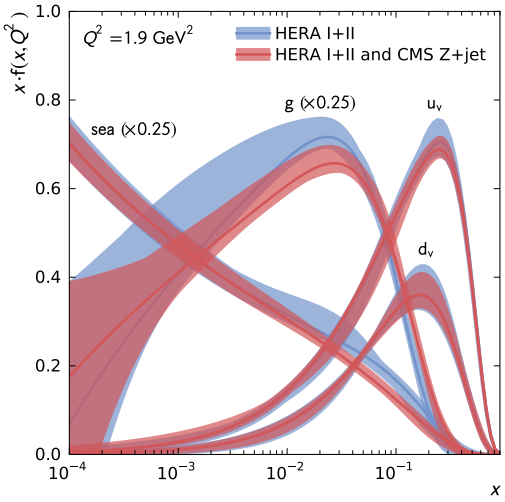
The jet calibration effort comprises the work of several analysis groups within the CMS Collaboration. For the data-driven calibration, the $\text{Z}(\rightarrow \mu^+\mu^-)$+jet analysis has provided the leading contribution. The uncertainties on the jet energy scale were greatly decreased for the data recorded at $\sqrt{s} = 8\,\text{TeV}$, see Figure 7.1(a). This has lead to significant improvements of the experimental precision in most physics analyses of the CMS Collaboration.

For studies of the proton structure, the $\text{Z}(\rightarrow \text{e}^+\text{e}^-)$+jet cross section was measured differentially as a function of transverse momentum and rapidity. This choice of variables allows to divide the phase space into regions corresponding to particular regions of proton momentum fraction x of the parton. The electron energy scale, reconstruction efficiencies, background process contributions and the unfolding procedure were carefully studied. The experimental and theoretical uncertainties were estimated. Theory calculations were performed at next-to-leading-order accuracy for different PDFs. By comparing the results of measurement and simulation, the parameters of the PDFs were fitted.

The predictions from perturbative QCD were confirmed for most of the studied phase space. The experimental uncertainties were found to be at the level of only a few percent. Including $\text{Z}(\rightarrow \text{e}^+\text{e}^-)$+jet data for PDF determination decreases PDF uncertainties and leads to a harder gluon, see Figure 7.1(b). Theory predictions will benefit from PDF improvements as were presented in this thesis. Opportunities for further improvements of the analysis are dedicated studies of the electron reconstruction at higher rapidities and a theory prediction of the $\text{Z}(\rightarrow \text{e}^+\text{e}^-)$+jet cross section at next-to-next-to-leading-order accuracy.



(a) Evolution of jet energy scale uncertainties in the 8 TeV data. The yellow band corresponds to the best-case scenario of a Z+jet flavour mixture. The improvement originates mainly from refined analysis methods and more data for calibration. The data-driven calibration with $Z(\rightarrow \mu^+\mu^-)$+jet events has made a significant contribution to these results. Adapted from [109, 111].



(b) Effects of $Z(\rightarrow e^+e^-)$+jet data on PDF determination. The PDF fits are based on deep-inelastic scattering data from the HERA collider. The inclusion of $Z(\rightarrow e^+e^-)$+jet data leads to a harder gluon. Additionally, the PDF uncertainties are reduced.

Figure 7.1: Results of Z+jet precision studies presented in this thesis.

Precision measurements of Z+jet events have lead to significant uncertainty reductions for the jet energy scale in the CMS experiment and the PDFs. The obtained results can not only improve many LHC physics analyses but they also contribute to our understanding of nature.

The analysis of vast amounts of data with a short turnaround time posed an unprecedented challenge. It was solved with the collaborative development of a high-performance software framework for data analysis. The enormous computational requirements were fulfilled with the pioneering use of cloud computing and coordinated caching systems. These technological accomplishments have paved the way for future studies with even higher computational demands.

In 2015, the LHC resumed operation at an increased centre-of-mass energy of 13 TeV. With the larger accessible phase space, even higher precision can be achieved as soon as enough data are collected. Additional refinements of reconstruction methods, theory predictions and analysis procedures may also lead to improvements. For future precision studies, the results presented in this thesis constitute an excellent foundation.

A Appendix

A.1 Technical Information

In this section, technical details of the $Z(\rightarrow \mu^+\mu^-)$+jet and $Z(\rightarrow e^+e^-)$+jet analyses are listed.

Data certification is implemented via a list of certified run periods:
`Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt`.
This file is provided by the CMS Collaboration and read in by ARTUS. The certified data corresponds to $19.712\,\text{fb}^{-1}$.

Jet energy correction version and the corresponding uncertainty is called WINTER 14. The correction files are provided by the CMS Collaboration in the format of TXT files.

Names and sizes of simulated and data samples are given in Table A.1.

Trigger paths

- $Z(\rightarrow \mu^+\mu^-)$+jet analysis: `HLT_Mu17_Mu8`
- $Z(\rightarrow e^+e^-)$+jet analysis: `HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL-_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL`

Jet Identification

For jets reconstructed from Particle-Flow candidates, the following selection is recommended by the CMS Collaboration following the *tight* identification criteria [148]: For all jets:

- Neutral hadron energy fraction <0.99
- Neutral EM (electron/photon) energy fraction <0.99
- Number of particle constituents >1
- Muon energy fraction <0.8

For jets reconstructed within tracker coverage ($|\eta| < 2.4$):

- Charged hadron energy fraction >0

- Number of charged constituents >0
- Charged EM (electron/photon) energy fraction <0.99

Muon Identification

A series of quality criteria are applied to the muons according to CMS recommendation for "tight" muons, i.e. identification criteria resulting in a high purity of the collection of reconstructed muons [149]. Only muons which fullfill the following requirements are selected:

- The muon candiate is a *global muon*, i.e. reconstructed from hits in the muon chambers and matching hits in the tracking system.
- The muon candidate has been labelled a valid muon by the Particle-Flow algorithm.
- $\chi^2/n_{\text{d.o.f.}}$ of the global-muon track fit is less than ten.
- At least one muon chamber hit is included in the global-muon track fit.
- Muon segments in at least two stations of the muon system are involved.
- The track in the inner tracker has a transverse impact parameter $d_{\text{xy}} < 2\,\text{mm}$ with respect to the primary vertex.
- The longitudinal distance of the track in the inner tracker with respect to the primary vertex is $d_{\text{z}} < 5\,\text{mm}$.
- The number of pixel hits is greater than zero.
- The number of tracker layers with hits is greater than five.

Electron Identification

Several quality criteria are applied to reconstructed electrons, see Table A.2. This follows the recommendation of the CMS Collaboration for "tight" electron identification, see [150].

Table A.1: List of used data sets.

Data set name	Nr. of events
Data	
/DoubleMu/Run2012A-22Jan2013-v1/AOD	5 636 274
/DoubleMuParked/Run2012B-22Jan2013-v1/AOD	29 308 627
/DoubleMuParked/Run2012C-22Jan2013-v1/AOD	36 820 243
/DoubleMuParked/Run2012D-22Jan2013-v1/AOD	38 006 513
/DoubleElectron/Run2012A-22Jan2013-v1/AOD	12 964 286
/DoubleElectron/Run2012B-22Jan2013-v1/AOD	23 571 931
/DoubleElectron/Run2012C-22Jan2013-v1/AOD	33 843 769
/DoubleElectron/Run2012D-22Jan2013-v1/AOD	34 526 899
Simulated signal	
/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball/[RD]	30 459 503
/DYToEE_M-20_CT10_TuneZ2star_8TeV-powheg-pythia6/[S10v2]	3 297 045
Simulated backgrounds	
/WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball/[S10v2]	57 709 905
/TTJets_FullLeptMGDecays_8TeV-madgraph/[RD]	21 675 970
/T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/[S10v1]	497 658
/Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/[S10v1]	493 460
/DYToTauTau_M-20_CT10_TuneZ2star_8TeV-powheg-pythia6/[S10v1]	3 295 238
/WW_TuneZ2star_8TeV_pythia6_tauola/[S10v1]	10 000 431
/WZ_TuneZ2star_8TeV_pythia6_tauola/[S10v1]	10 000 283
/ZZ_TuneZ2star_8TeV_pythia6_tauola/[S10v1]	9 799 908
/QCD_Pt_20_30_EMEnriched_TuneZ2star_8TeV_pythia6/[S10v1]	35 040 695
/QCD_Pt_30_80_EMEnriched_TuneZ2star_8TeV_pythia6/[S10v1]	33 088 888
/QCD_Pt_80_170_EMEnriched_TuneZ2star_8TeV_pythia6/[S10v1]	34 542 763
/QCD_Pt_170_250_EMEnriched_TuneZ2star_8TeV_pythia6/[S10v1]	31 697 066
/QCD_Pt_250_350_EMEnriched_TuneZ2star_8TeV_pythia6/[S10v1]	34 611 322
/QCD_Pt_350_EMEnriched_TuneZ2star_8TeV_pythia6/[S10v1]	34 080 562

[RD] Summer12_DR53X-PU_RD1_START53_V7N-v1/AODSIM
[S10v1] Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
[S10v2] Summer12_DR53X-PU_S10_START53_V7A-v2/AODSIM

Table A.2: Electron identification criteria.

Quantity	Maximum value Barrel	Endcap
$\lvert\Delta\eta(\mathrm{SC}, \mathrm{track})\rvert$	0.004	0.005
$\lvert\Delta\phi(\mathrm{SC}, \mathrm{track})\rvert$	0.03	0.02
SC shape covariance (*SigmaIEtaIEta*)	0.01	0.03
Ratio of HCAL and ECAL energy (*hadronicOverEm*)	0.12	0.1
Missing hits in inner tracker $\leq$	0	
Distance between track and primary vertex in x-y-plane	0.02	
Distance between track and primary vertex in z-direction	0.1	
$\lvert(1/(E_{\mathrm{ECAL}}) - 1/(p_{\mathrm{track}})\rvert$	0.05	
Particle-Flow isolation ($\Delta R < 0.3$)	0.18	

A.2 Additional Analysis Results

A.2.1 Z($\rightarrow \mu^+\mu^-$)+Jet Calibration Analysis

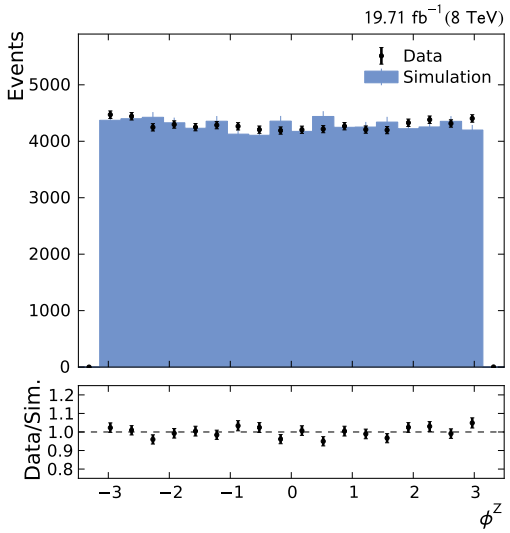


Figure A.1: Z boson ϕ distribution.

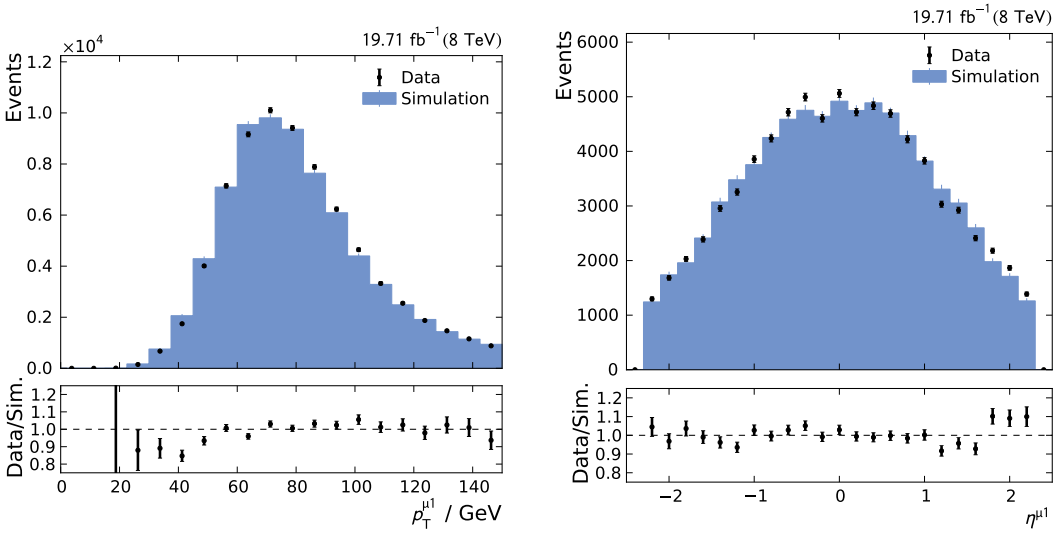


Figure A.2: Leading muon p_T (left) and η (right).

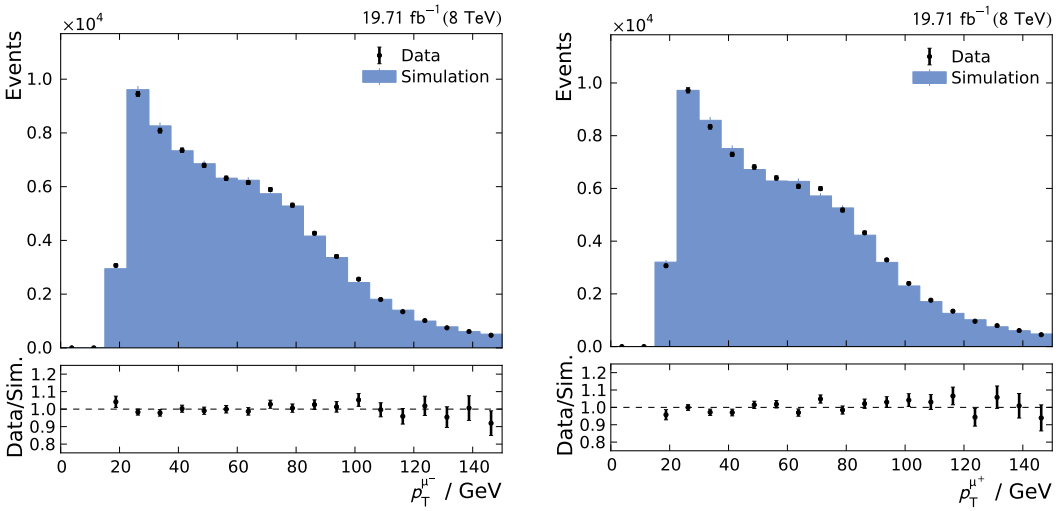


Figure A.3: μ^- (left) and μ^+ p_T (right).

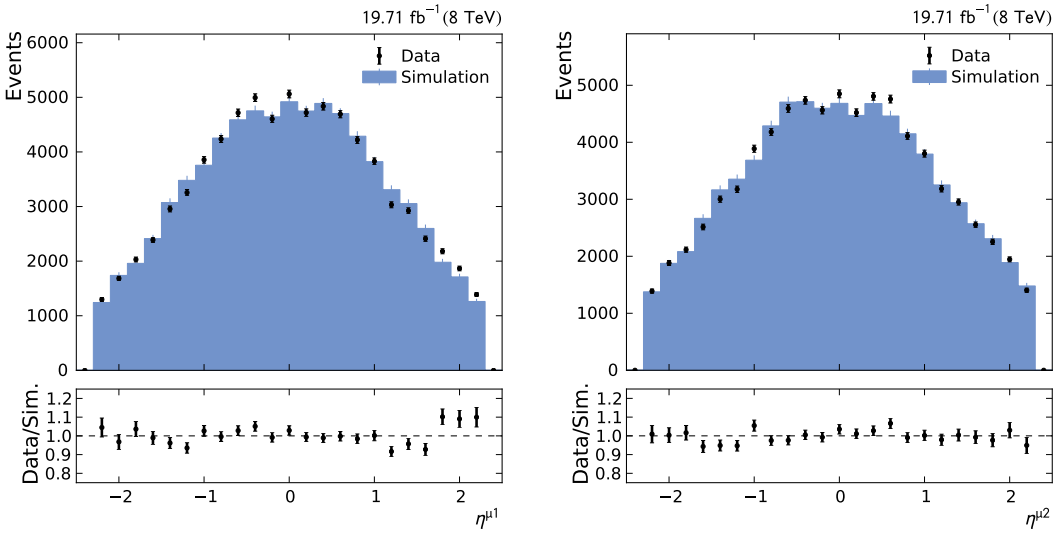


Figure A.4: Leading (left) and subleading (right) muon η.

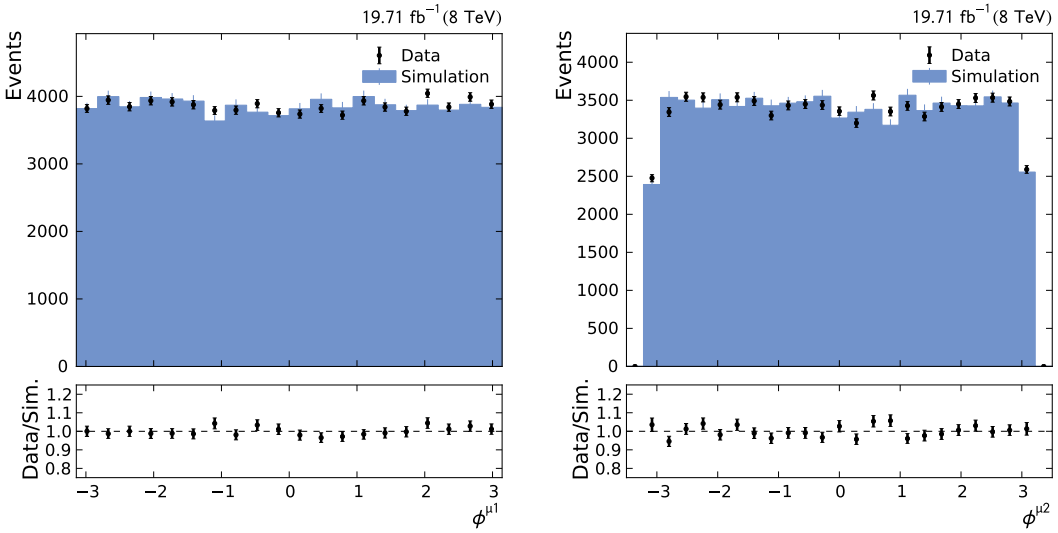


Figure A.5: Leading (left) and subleading (right) muon ϕ.

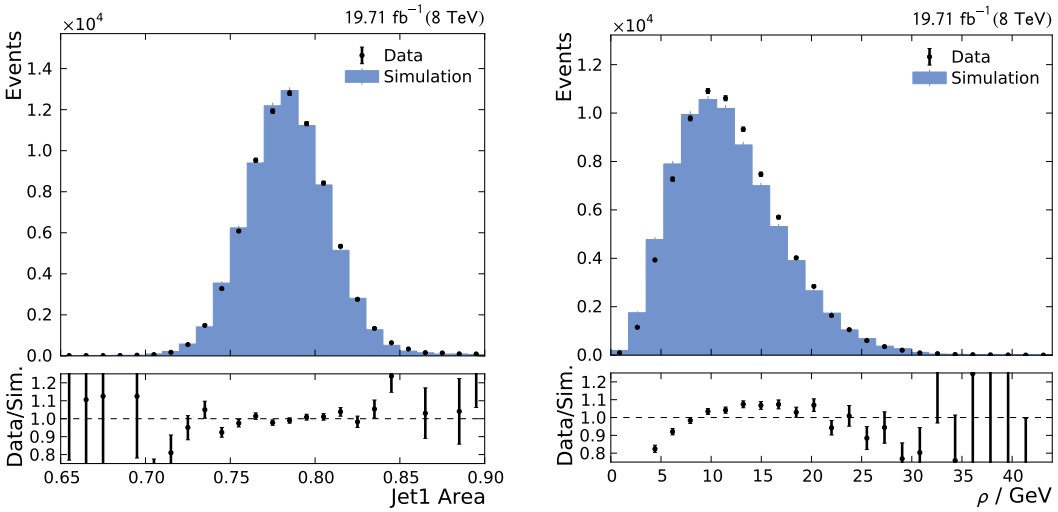


Figure A.6: Leading jet area (left) and event energy density ρ (right).

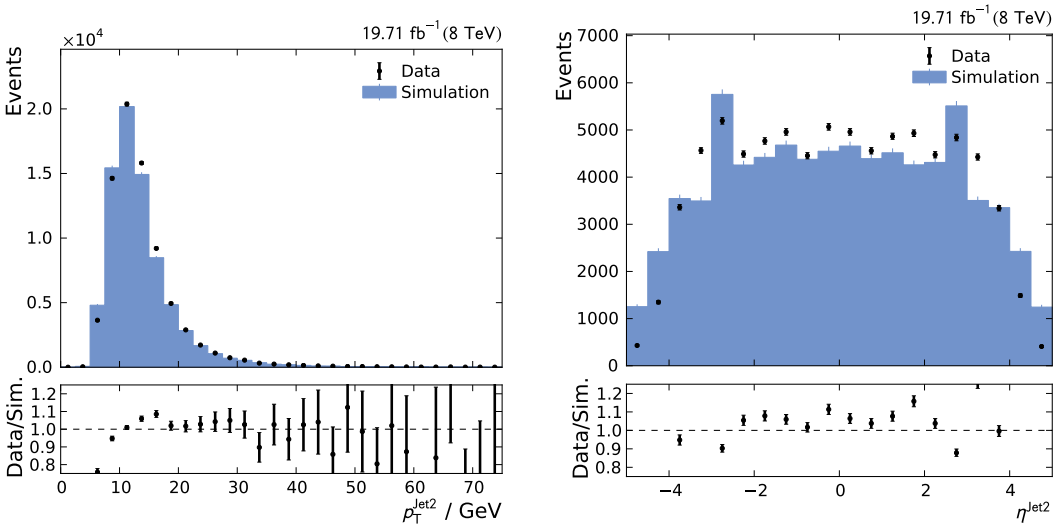


Figure A.7: Second jet p_{T} (left) and η (right).

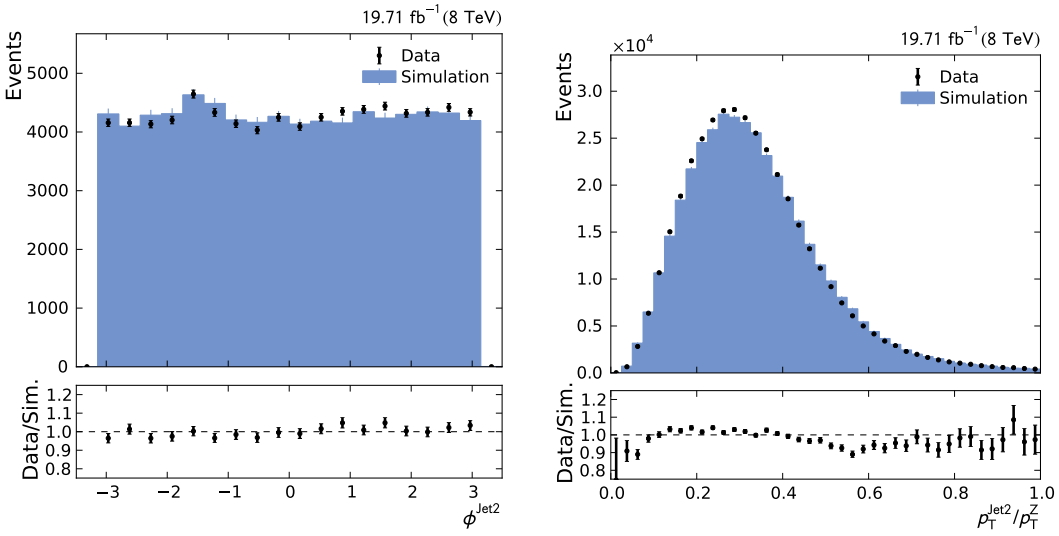


Figure A.8: Second jet ϕ (left) and $p_{\mathrm{T}}^{\mathrm{Jet2}}/p_{\mathrm{T}}^{\mathrm{Z}}$ (right).

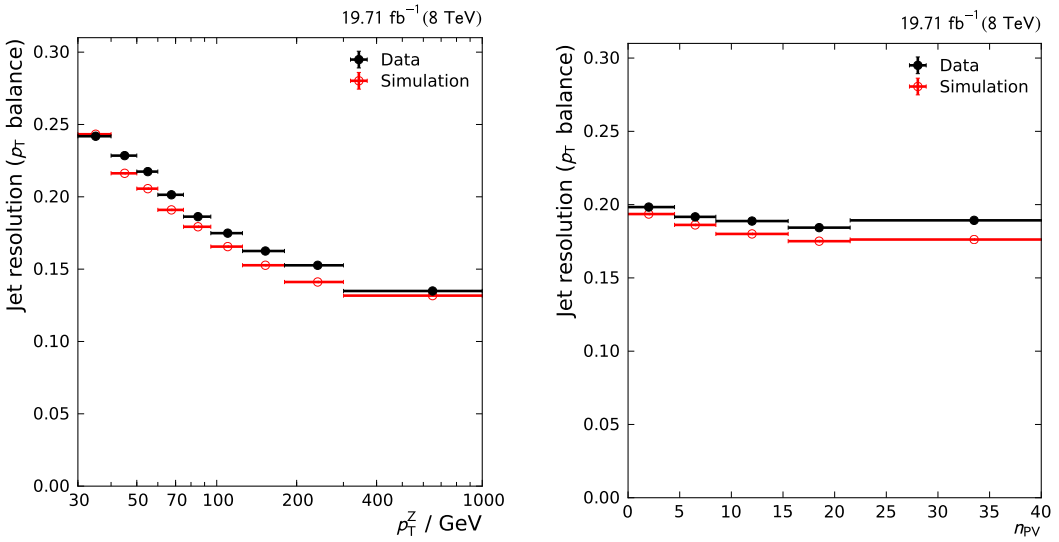


Figure A.9: Jet p_{T} resolution measured with the p_{T} balance method as a function ot Z boson p_{T} (left) and number of reconstructed primary vertices (right).

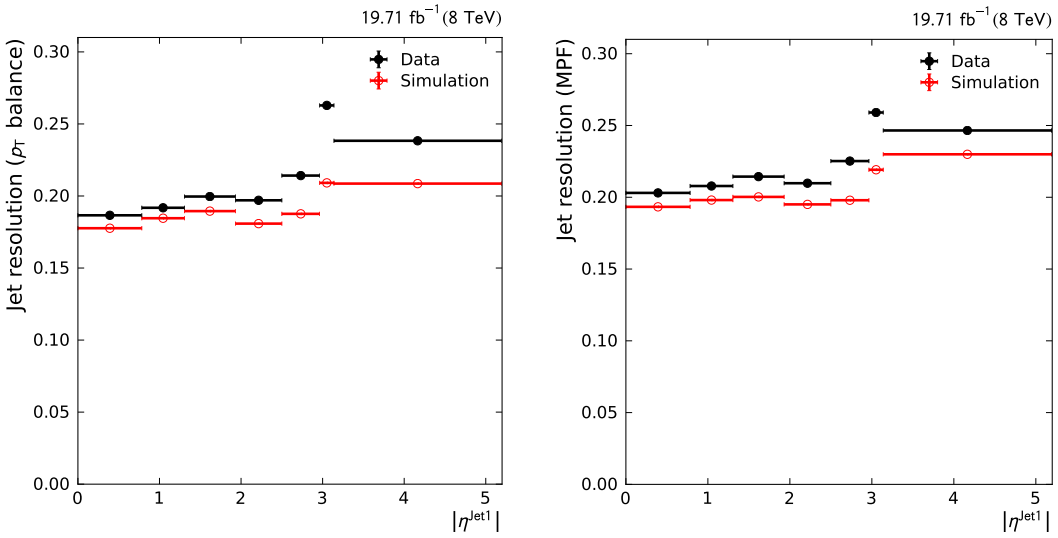


Figure A.10: Jet p_{T} resolution as a function of jet $|\eta|$ determined with the p_{T} balance (left) and the MPF response (right).

A.2.2 Z(→ e$^+$e$^-$)+Jet PDF Determination Analysis

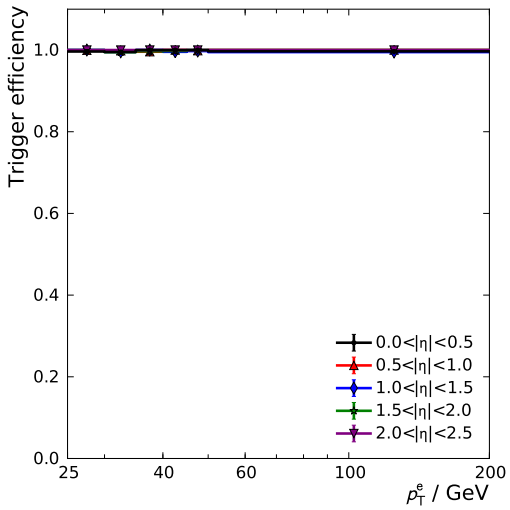


Figure A.11: Electron trigger efficiency. Because of the tight identification applied, the trigger efficiency is close to one for all pseudorapidity regions.

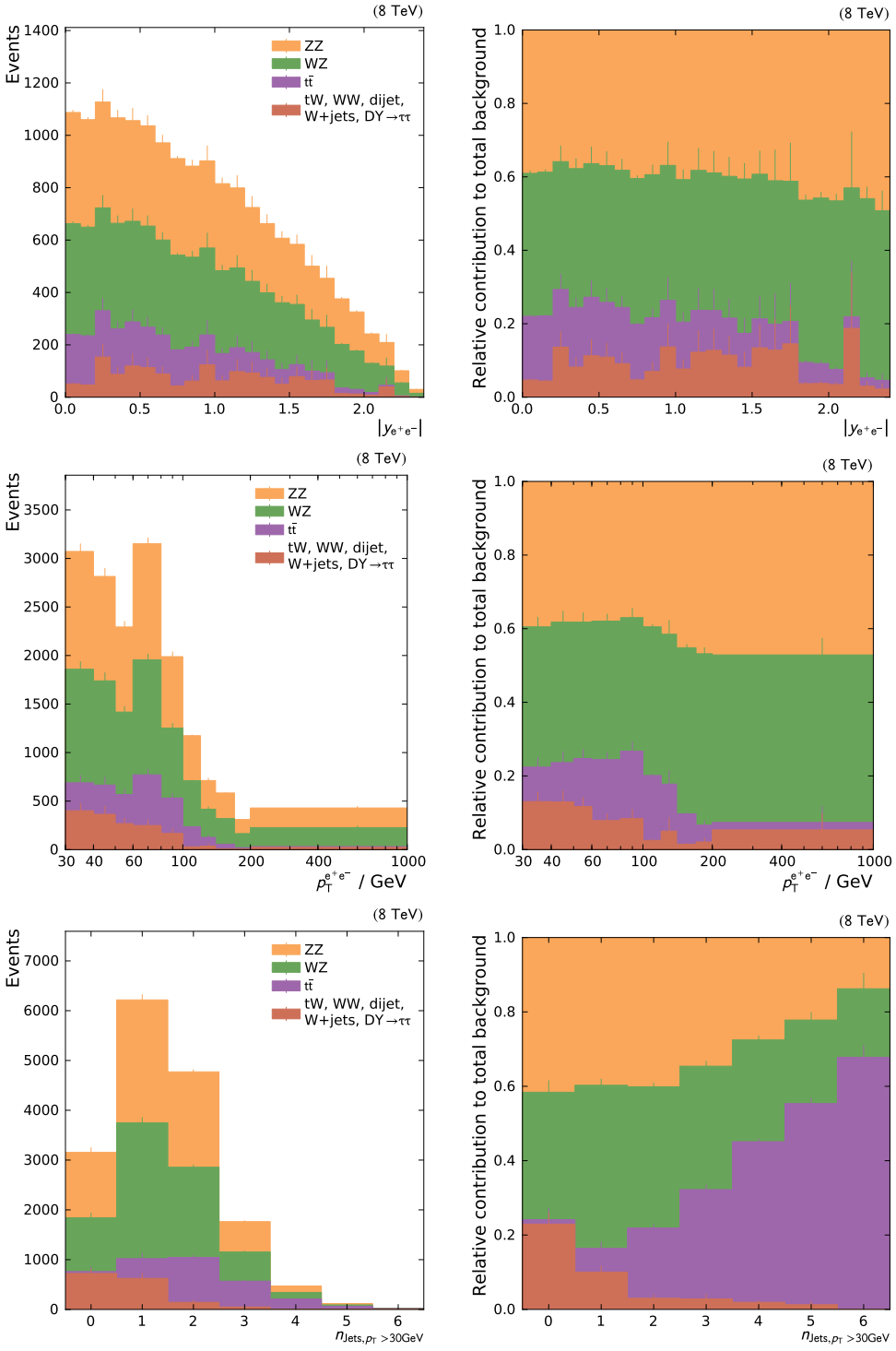


Figure A.12: Backgrounds contributions as a function of dielectron rapidity (top row), dielectron p_{T} (middle row) and number of jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$. The left column shows the absolute expectation, the right column the relative contribution of each background category per bin.

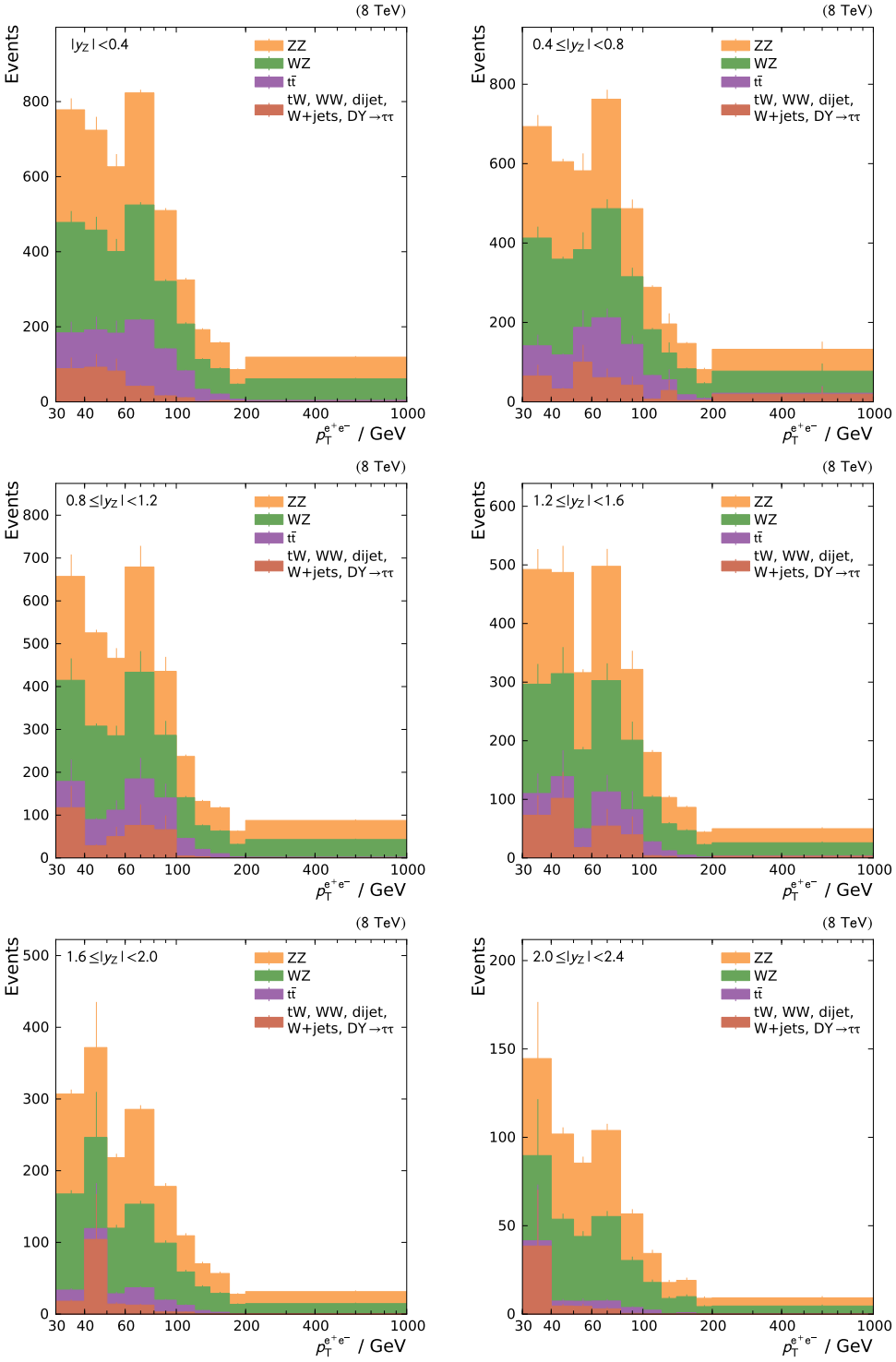


Figure A.13: Background contributions for the double-differential measurement.

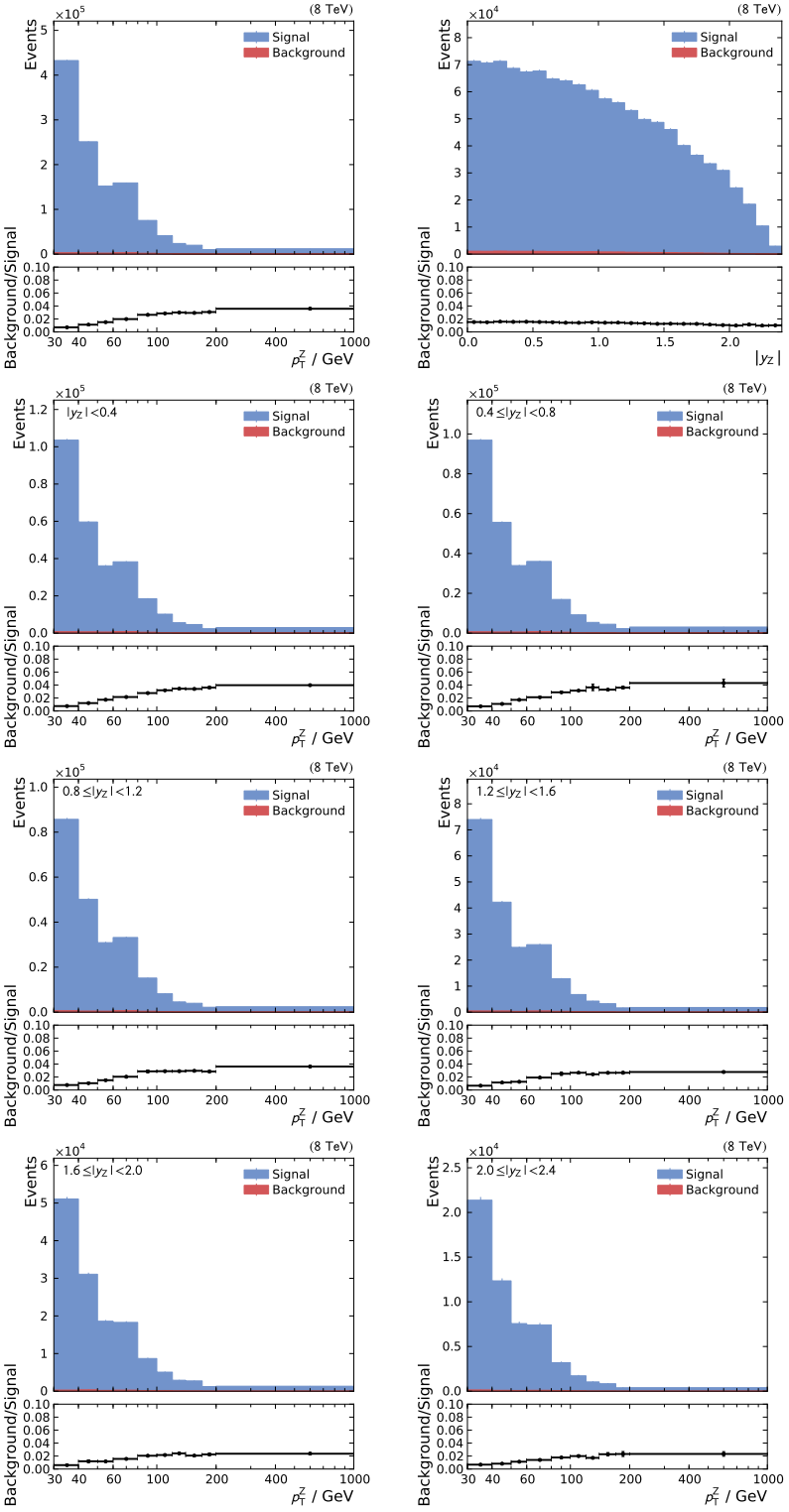
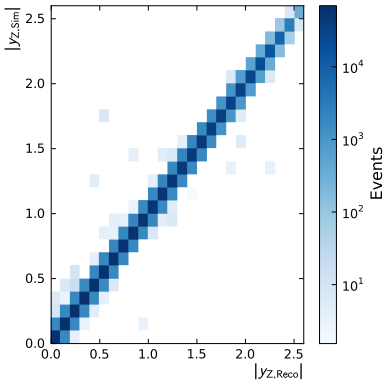
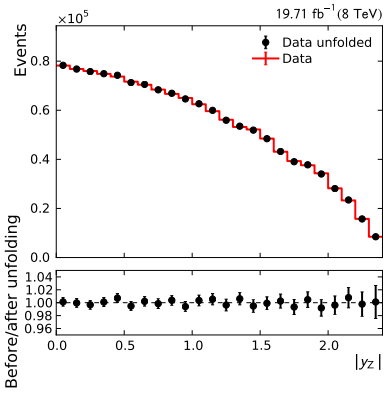


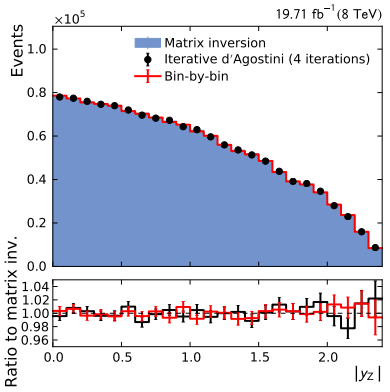
Figure A.14: Background-signal ration for the double-differential measurement.



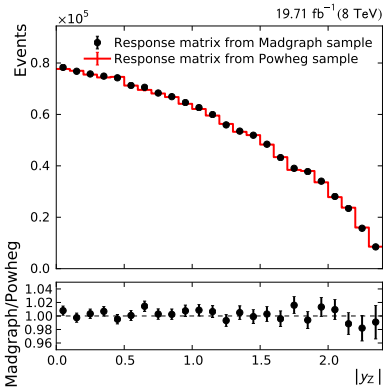
(a) Response matrix



(b) Comparison of distribution before and after unfolding



(c) Result of unfolding with different methods



(d) Result of unfolding with response matrices from different samples

Figure A.15: Unfolding of the rapidity distribution.

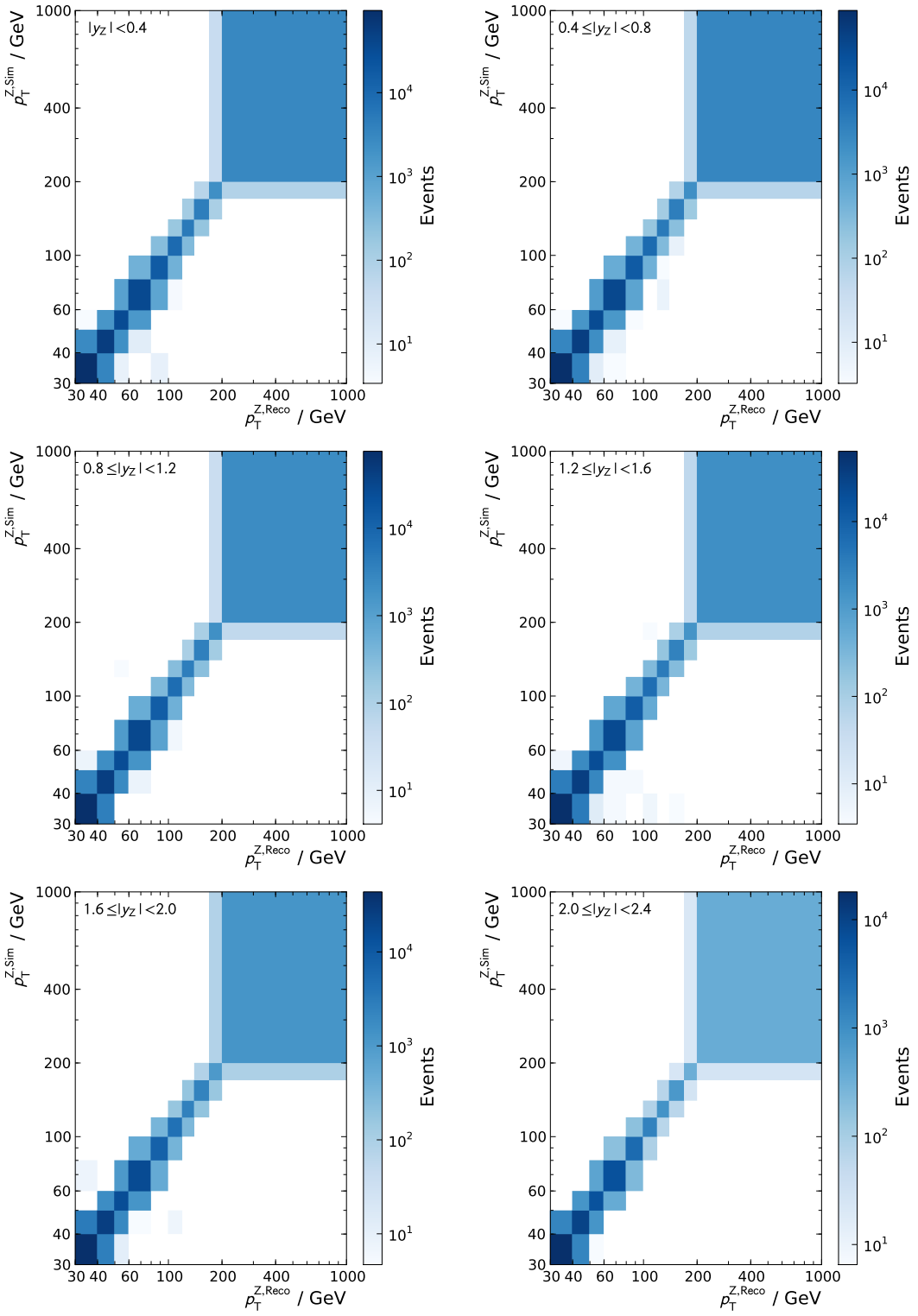


Figure A.16: Response matrices used for the unfolding.

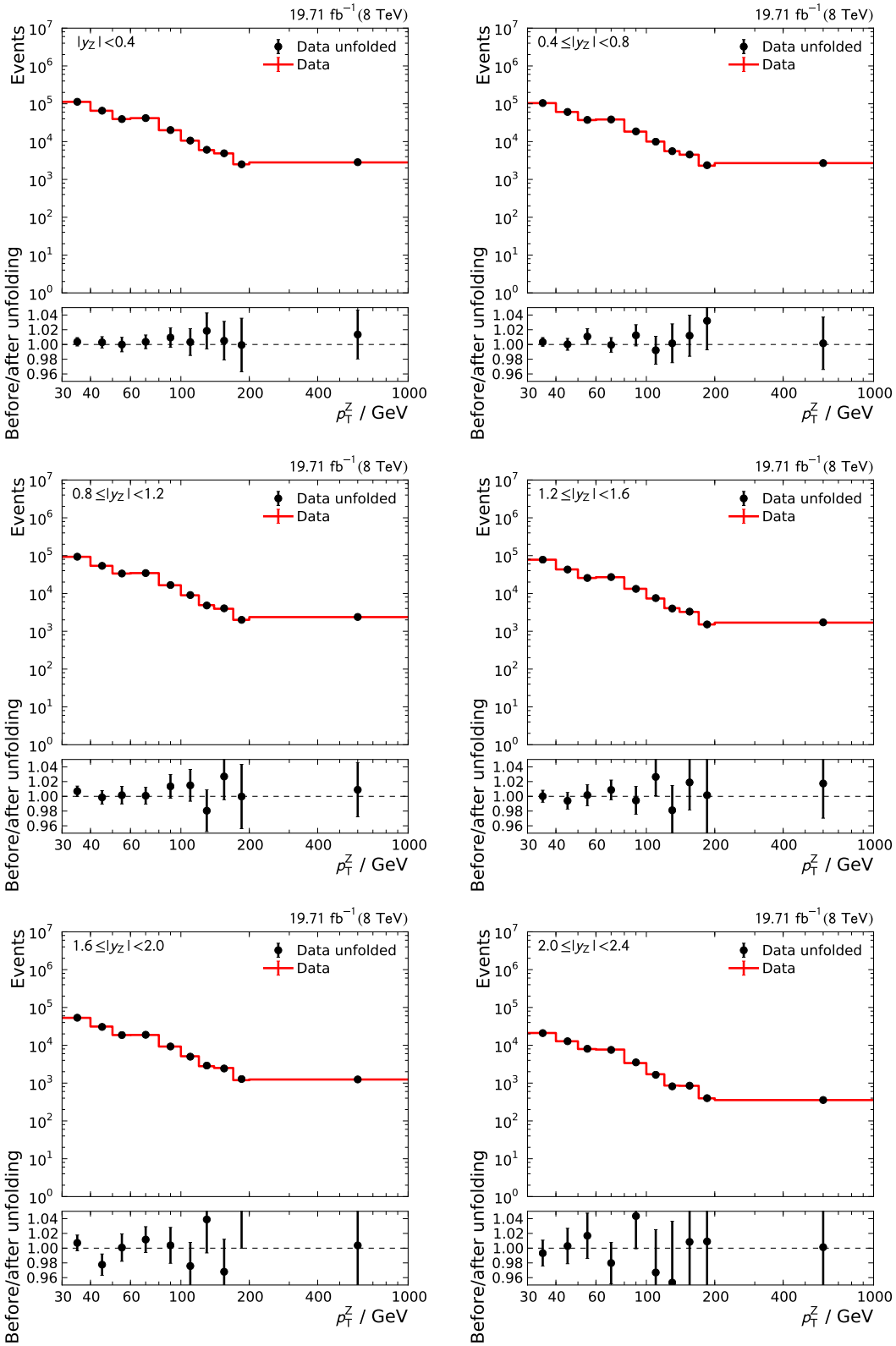


Figure A.17: Comparison of cross section before and after unfoldong.

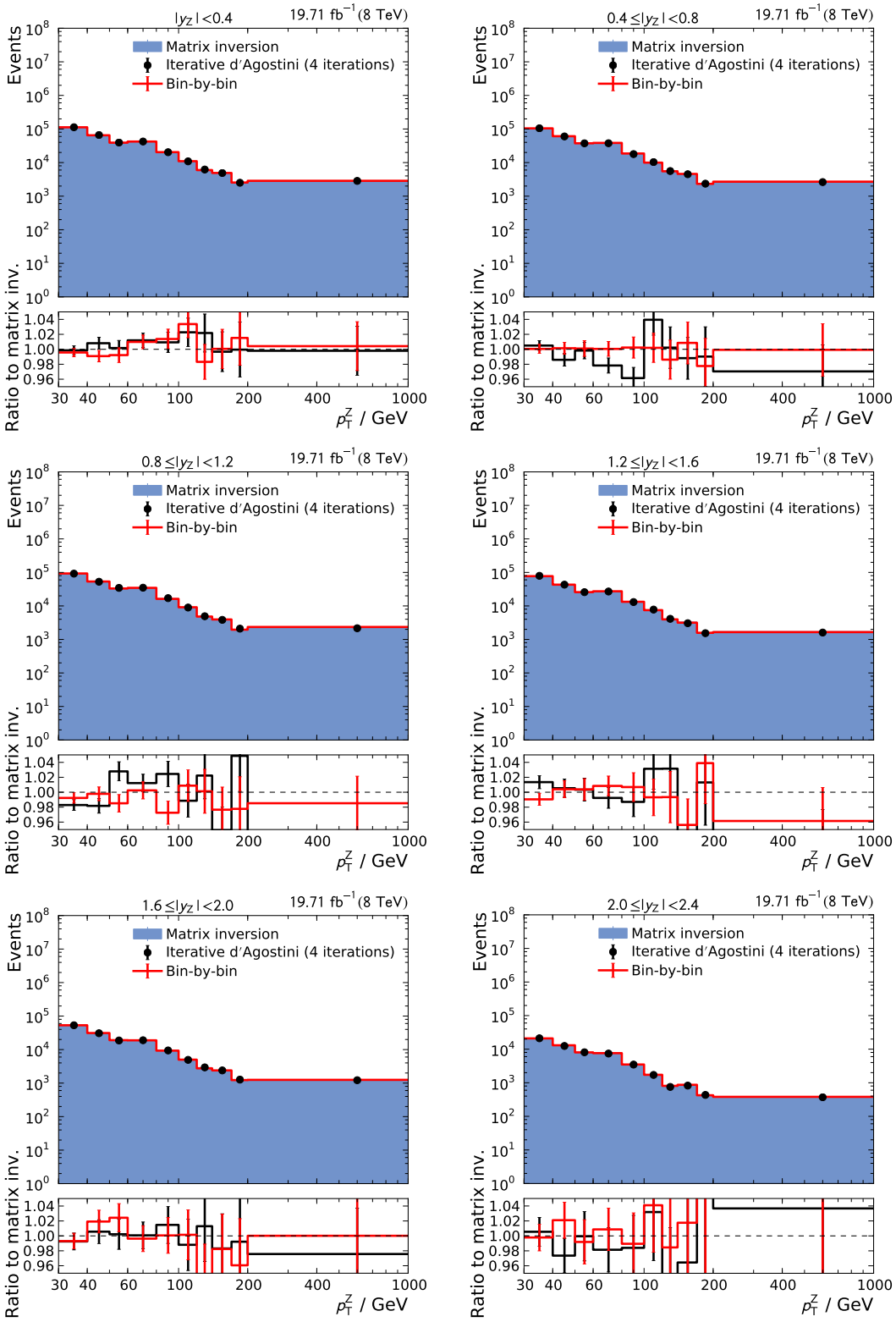


Figure A.18: Comparison of different unfolding methods.

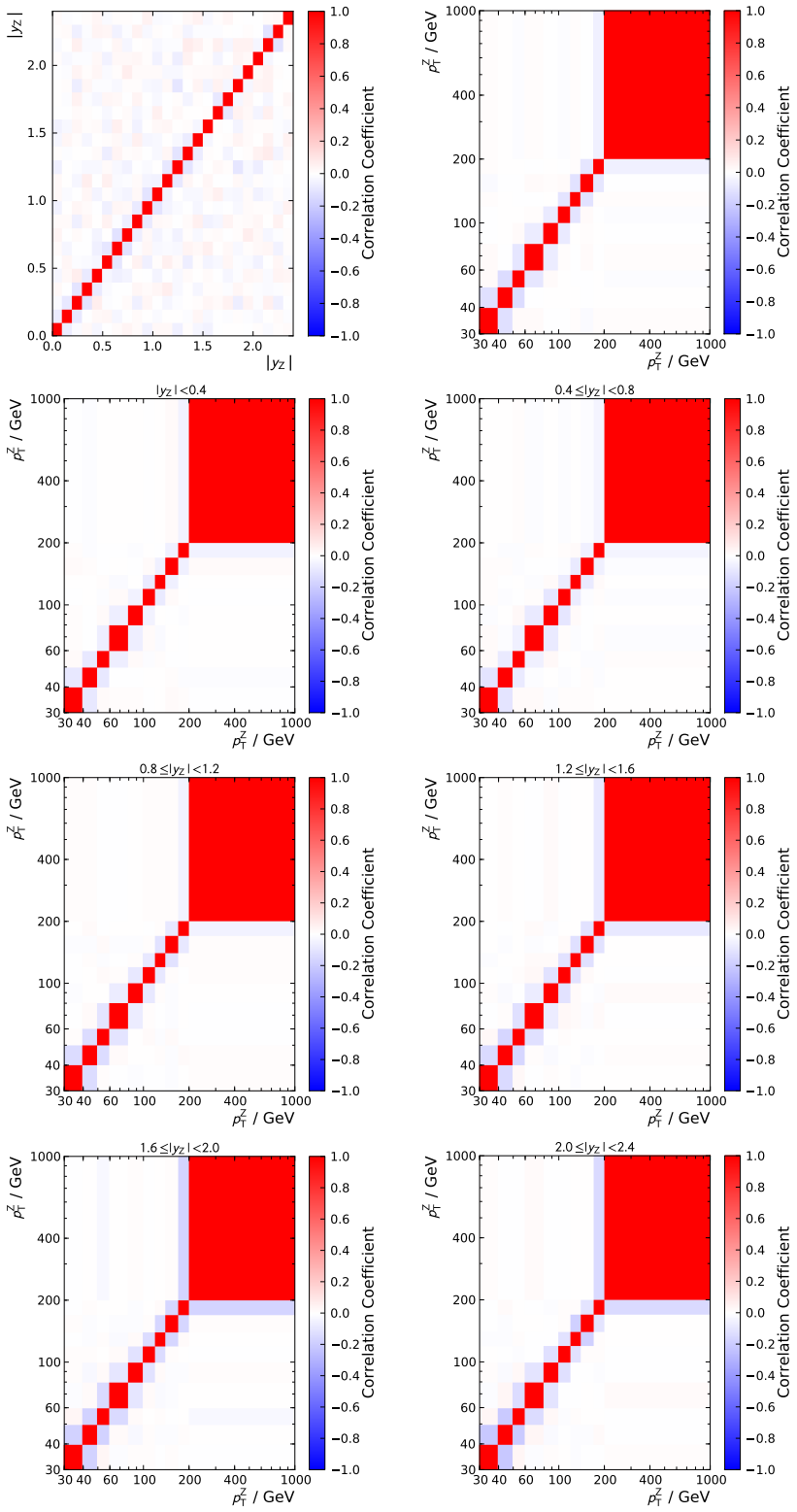


Figure A.19: Statistical correlation matrices. The correlation is introduced by the Toy Monte Carlo method used for the unfolding.

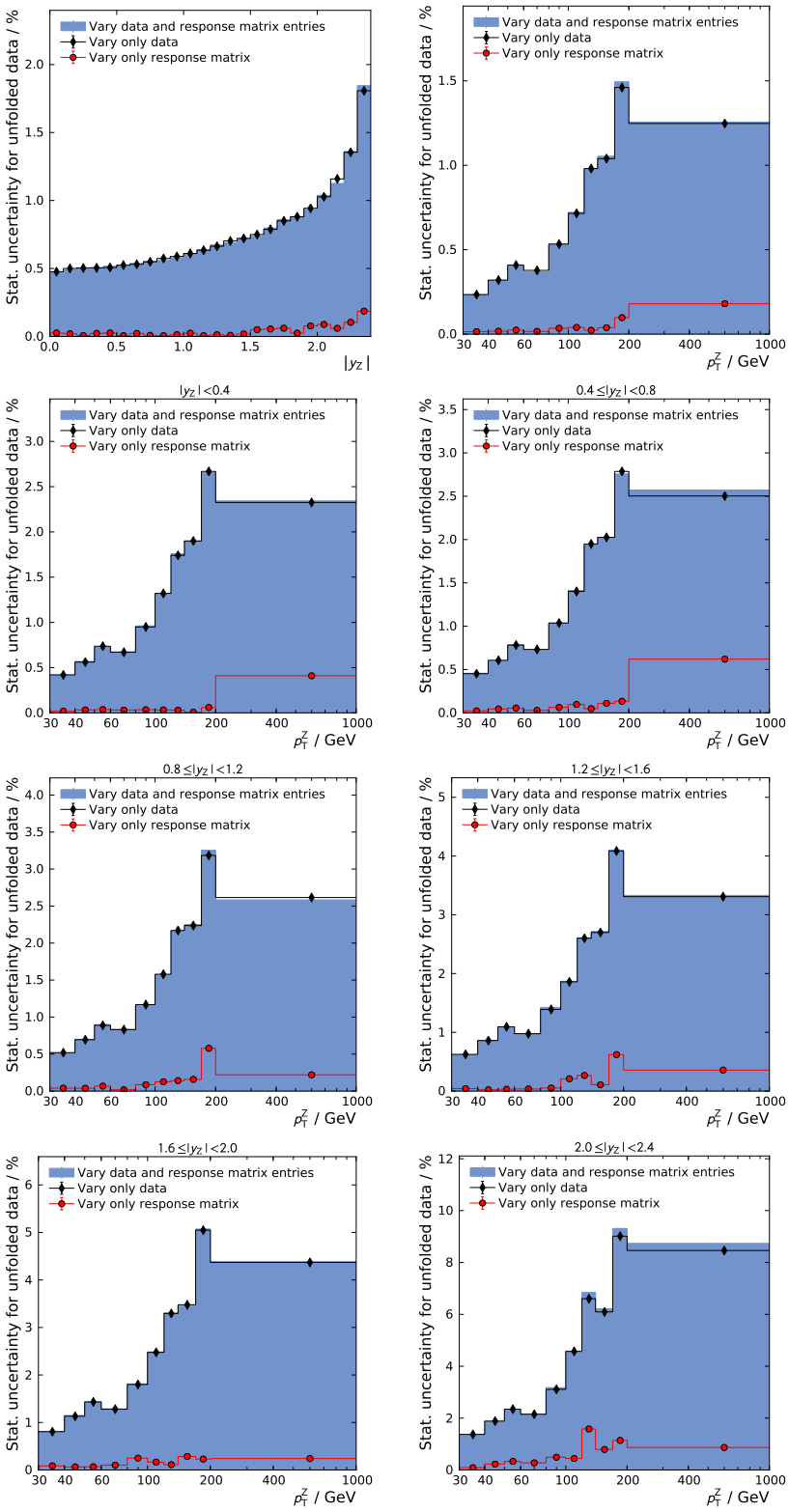


Figure A.20: Statistical uncertainty, calculated via pseudo-experiments from the statistical uncertainty in the data and the response matrix.

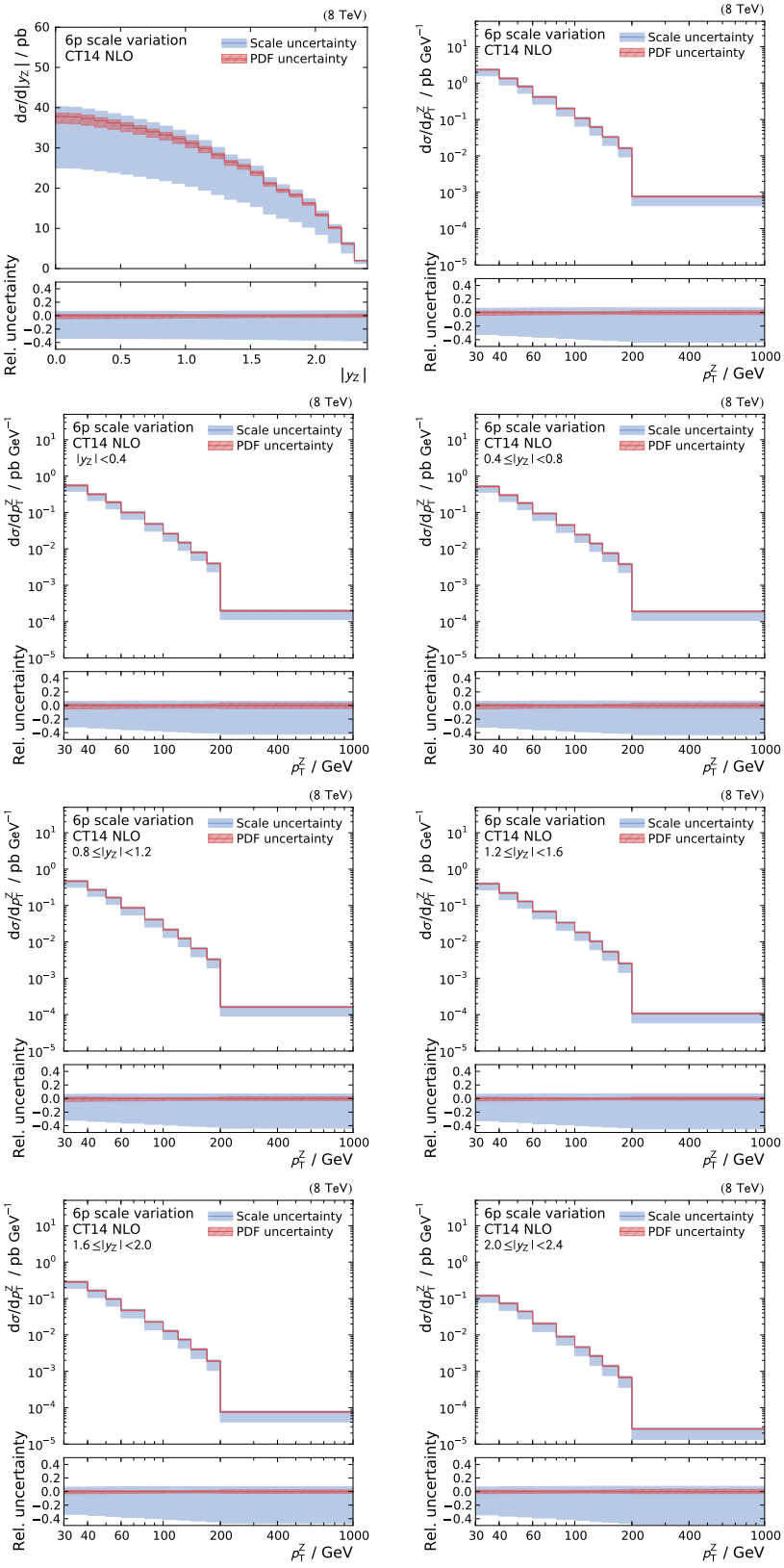


Figure A.21: Theory prediction with scale and PDF uncertainties.

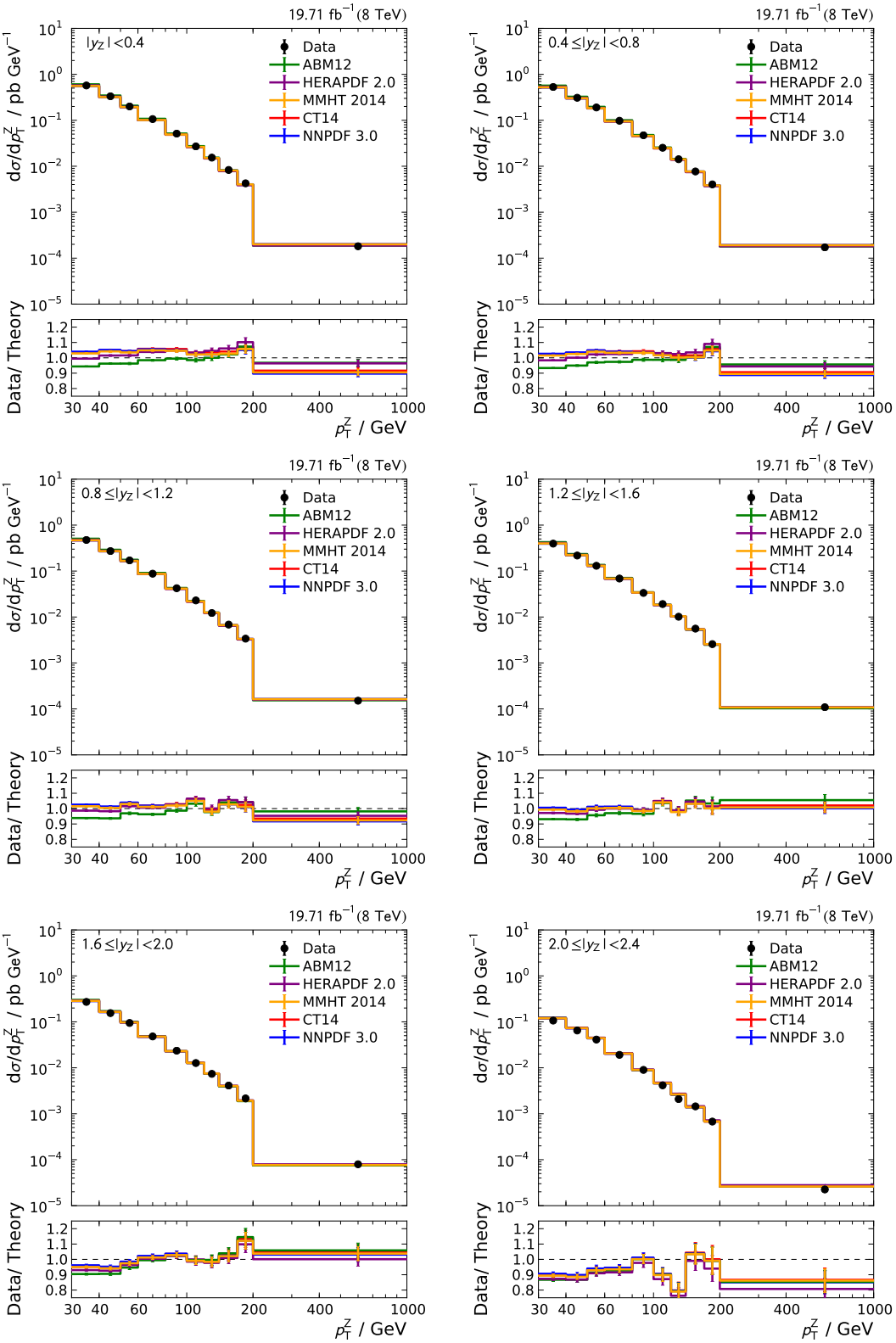


Figure A.22: Comparison of data with predictions based on different PDF sets.

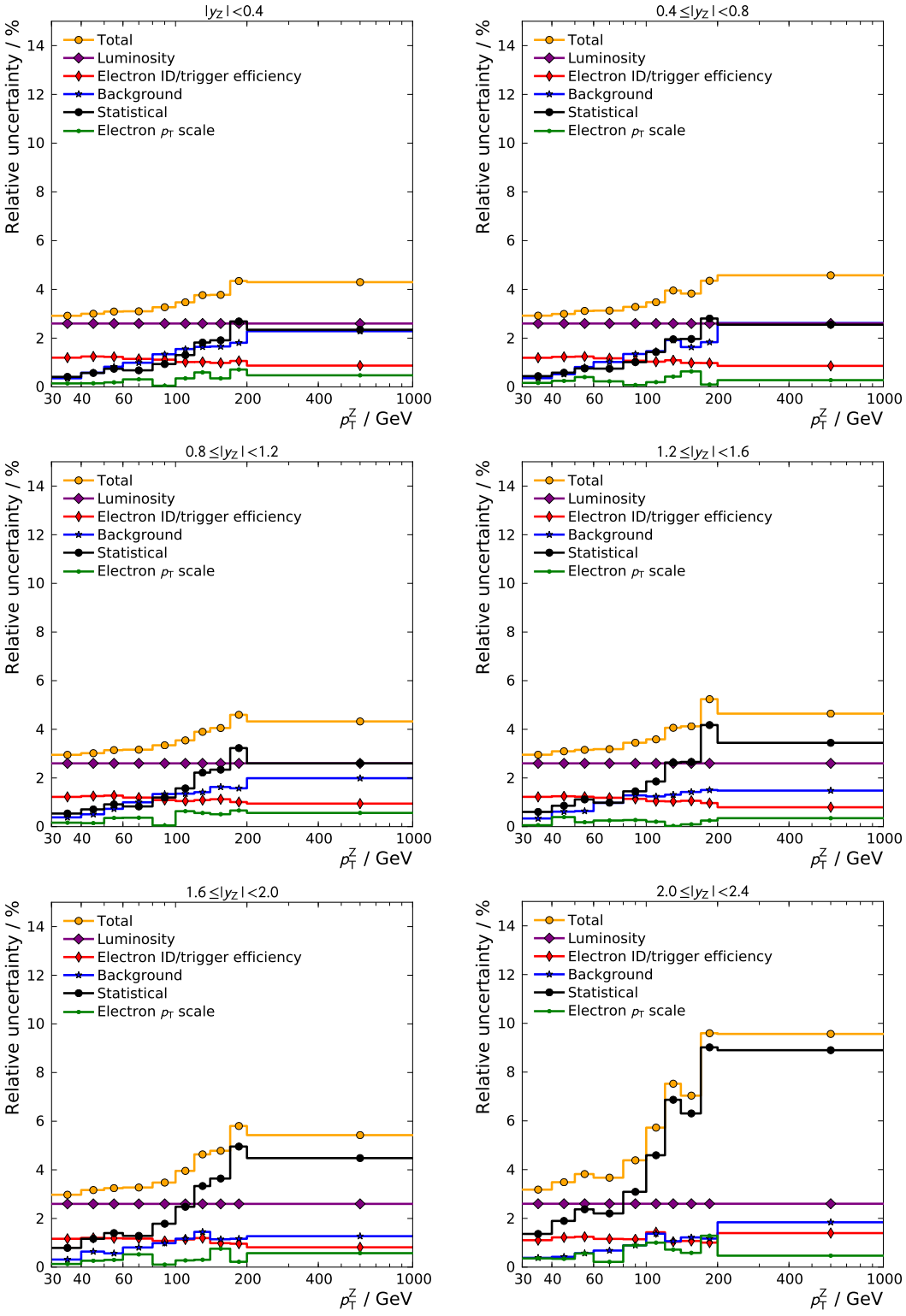


Figure A.23: Experimental uncertainties as a function of $p_{\mathrm{T}}^{\mathrm{Z}}$ for different regions of $|y_{\mathrm{Z}}|$.

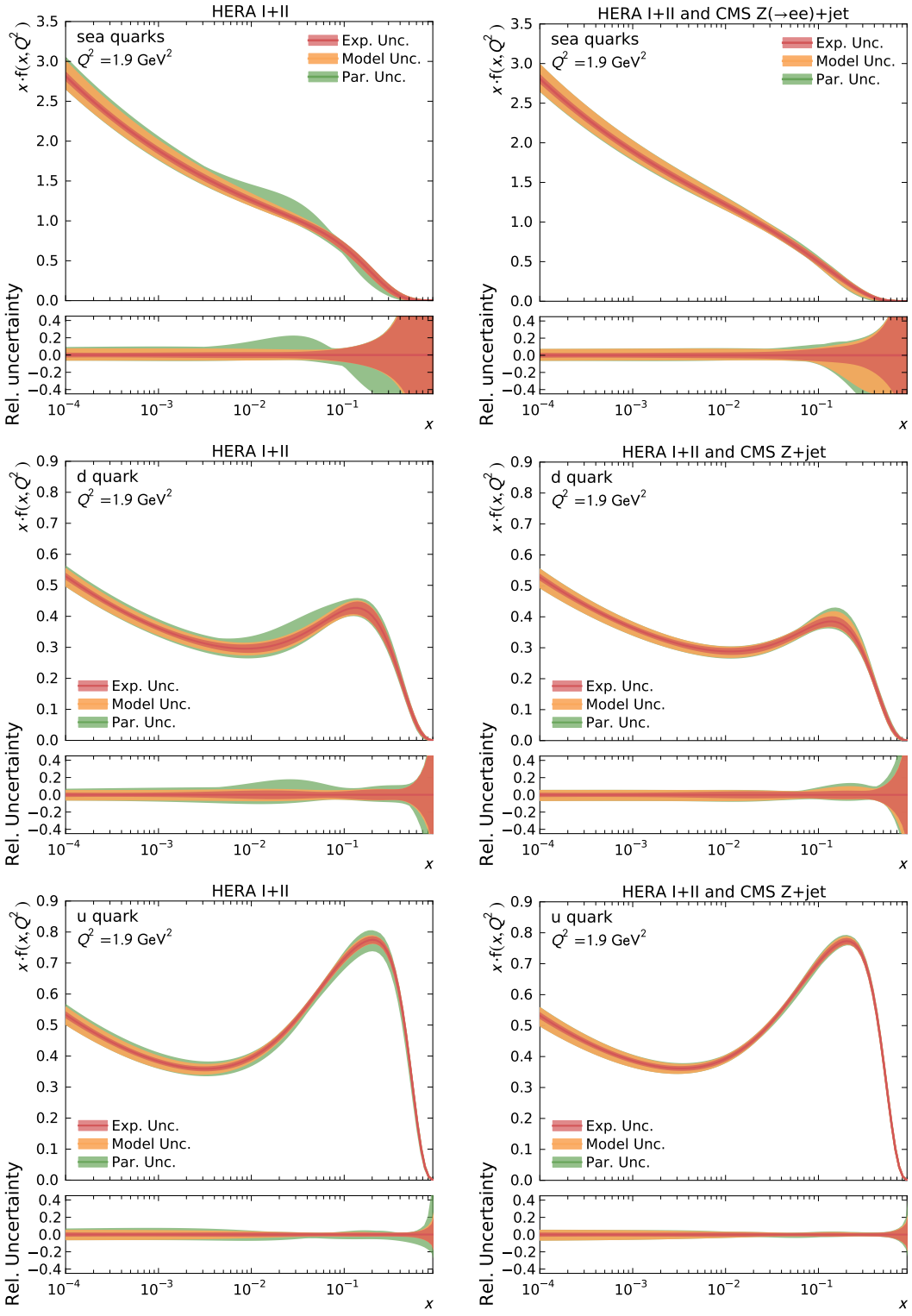


Figure A.24: PDFs at $Q^2 = 1.9\,\mathrm{GeV}^2$ with uncertainties. Experimental, model and parametrization uncertainties are added in quadrature.

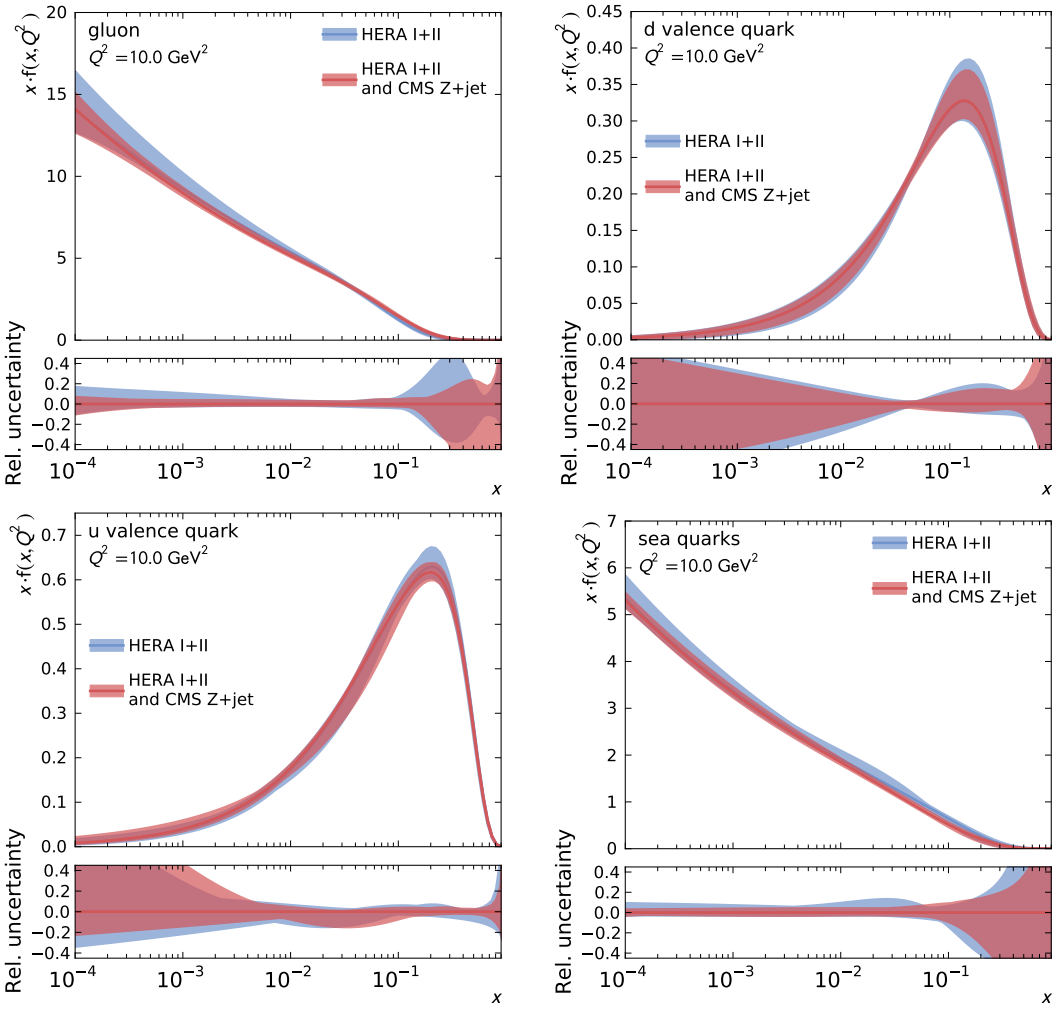


Figure A.25: Comparison of PDFs at $Q^2 = 10\,\mathrm{GeV}^2$.

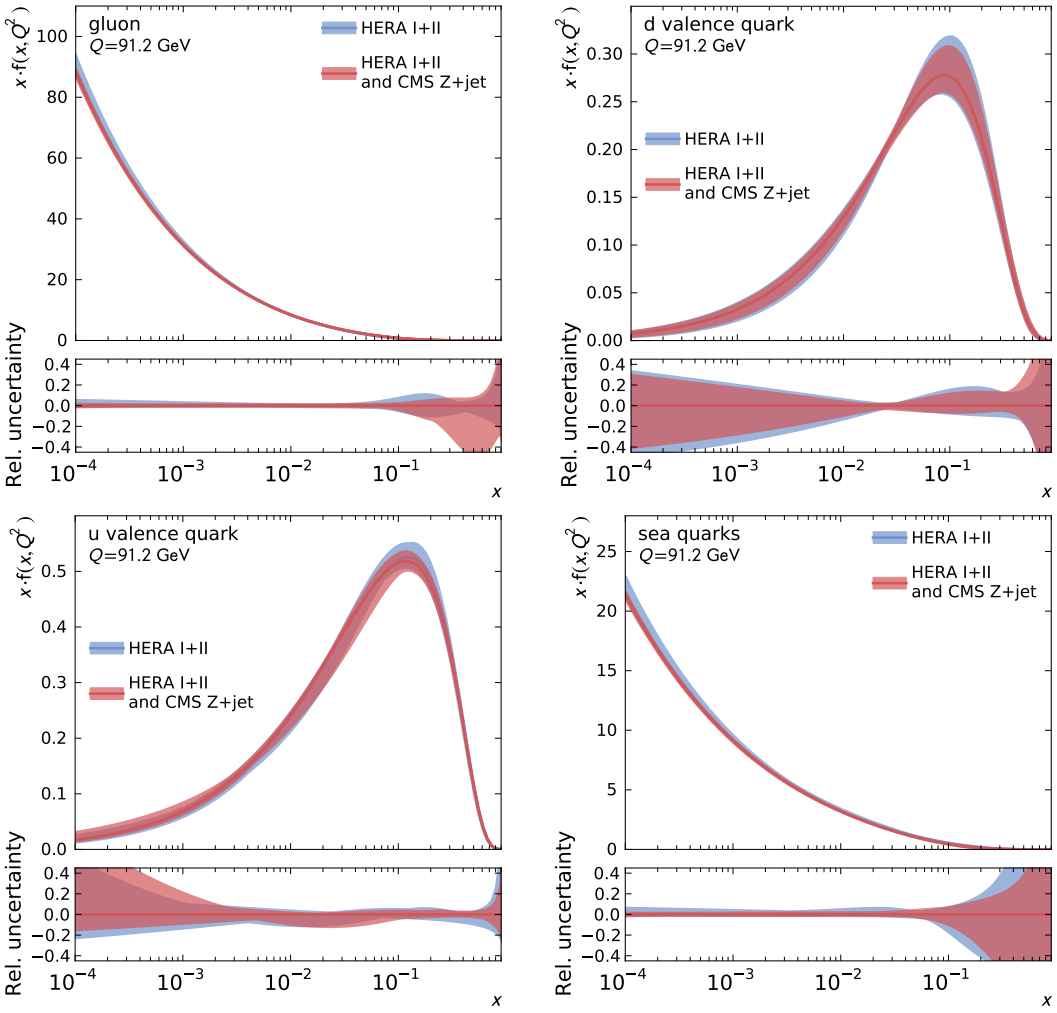


Figure A.26: Comparison of PDFs at $Q = m_Z$.

A.3 Software Configuration Details

The versions of the software used for the analyses detailed in this thesis are listed in Table A.3.

Table A.3: Software versions.

Software	Version
CMS data analysis	
CMSSW	5.3.14_patch2
ROOT	5.35
Kappa	2.0
Artus	8TeVFinal
Excalibur	8TeVFinal
Simulation and PDF fits	
Sherpa	2.2.0
BlackHat	0.9.9
Rivet	2.3.0
MCGrid	2.0
fastNLO	2.3.1pre
LHAPDF	6.1.5
xFitter	1.2.0

A.3.1 Sherpa Configuration File

The Sherpa configuration file used for the simulation of $Z(\rightarrow e^+e^-)$+jet events:

```
(run){
  EVENTS 1M;
  ANALYSIS Rivet;
  ANALYSIS_OUTPUT Rivet;

  HEPMC_USE_NAMED_WEIGHTS=1;
  HEPMC_EXTENDED_WEIGHTS=1;

  PDF_LIBRARY=CT10Sherpa;
  PDF_SET=ct10;
}(run);

(beam){
  % LHC collider setup
  BEAM_1 2212; BEAM_ENERGY_1 4000;
  BEAM_2 2212; BEAM_ENERGY_2 4000;
```

```
}(beam);

(integration){
  ERROR 0.01;
}(integration);

(me){
  ME_SIGNAL_GENERATOR BlackHat Comix Internal;
  EVENT_GENERATION_MODE Weighted;
  SCALES VAR{8315.18+PPerp2(p[2]+p[3])};
  RESULT_DIRECTORY Results_NLO;
}(me);

(model){
  MASS[6] = 1e20
}(model)

(processes){

  ### The Born, virtual, integrated and real-subtracted pieces
  Process 93 93 -> 11 -11 93;
  NLO_QCD_Mode Fixed_Order;
  NLO_QCD_Part B;
  Order (*,2);
  End process;

  Process 93 93 -> 11 -11 93;
  NLO_QCD_Mode Fixed_Order;
  NLO_QCD_Part I;
  Order (*,2);
  End process;

  Process 93 93 -> 11 -11 93;
  NLO_QCD_Mode Fixed_Order;
  NLO_QCD_Part V;
  Loop_Generator BlackHat;
  Order (*,2);
  End process;

  Process 93 93 -> 11 -11 93;
  NLO_QCD_Mode Fixed_Order;
  NLO_QCD_Part RS;
  Order (*,2);
  End process;
```

```
61
62 }(processes);
63
64
65 (selector){
66   Mass 11 -11 81. 101.
67   PTNLO 11 25 E_CMS
68   PTNLO -11 25 E_CMS
69   PseudoRapidityNLO 11 -2.4 2.4
70   PseudoRapidityNLO -11 -2.4 2.4
71   PT2NLO 11 -11 30 E_CMS
72 }(selector)
73
74
75 (analysis){
76   BEGIN_RIVET {
77     -a MCgrid_CMS_2015_Zee;
78     USE_HEPMC_SHORT 1;
79     USE_HEPMC_EXTENDED_WEIGHTS 1;
80     IGNOREBEAMS 1;
81   } END_RIVET;
82 }(analysis);
```

List of Figures

List of Tables

Bibliography

[1] B. Povh, *Teilchen und Kerne: Eine Einführung in die physikalischen Konzepte*, 8. Edition, Springer, 2009, ISBN: 978-3-540-68080-2.

[2] D. Griffiths, *Introduction to Elementary Particles*, John Wiley & Sons, 2008, ISBN: 978-3-527-61847-7.

[3] M. E. Peskin and D. V. Schroeder, *An Introduction To Quantum Field Theory (Frontiers in Physics)*, Westview Press, 1995, ISBN: 0201503972.

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

[5] 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** (2015) p. 186, DOI: 10.1140/epjc/s10052-015-3376-y, [arXiv:1412.1633].

[6] K. A. Olive et al., "Review of Particle Physics", *Chin. Phys.* **C38** (2014) p. 090001, DOI: 10.1088/1674-1137/38/9/090001.

[7] J. Rojo, *The Strong Interaction and LHC phenomenology*, University Lecture, 2014, URL: https://www2.physics.ox.ac.uk/sites/default/files/2014-03-31/qcdgrad_rojo_oxford_tt14_2_basics_pdf_40958.pdf.

[8] R. K. Ellis, W. J. Stirling and B. R. Webber, *QCD and Collider Physics*, vol. 8, Cambridge University Press, 1996, 1–435, ISBN: 978-0521545891.

[9] H. Abramowicz et al., "Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data", *Eur. Phys. J.* **C75** (2015) p. 580, DOI: 10.1140/epjc/s10052-015-3710-4, [arXiv:1506.06042].

[10] G. Sieber, "Bestimmung von Protonstruktur und starker Kopplung aus inklusiven Jet-Wirkungsquerschnitten am LHC", diploma thesis, KIT, 2013, IEKP-KA/2013-12.

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

[12] S. Alekhin, J. Blumlein and S. Moch, "The ABM parton distributions tuned to LHC data", *Phys. Rev.* **D89** (2014) p. 054028, DOI: 10.1103/PhysRevD.89.054028, [arXiv:1310.3059].

[13] S. Dulat et al., "New parton distribution functions from a global analysis of quantum chromodynamics", *Phys. Rev.* **D93** (2016) p. 033006, DOI: 10.1103/PhysRevD.93.033006, [arXiv:1506.07443].

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

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

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

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

[18] CMS Collaboration, *Standard Model Cross Sections for CMS at 8 TeV*, URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat8TeV (visited on 05/04/2016).

[19] J. Rojo et al., "The PDF4LHC report on PDFs and LHC data: Results from Run I and preparation for Run II", *J. Phys.* **G42** (2015) p. 103103, DOI: 10.1088/0954-3899/42/10/103103, [arXiv:1507.00556].

[20] S. Höche, "Introduction to parton-shower event generators", *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014) Boulder, Colorado, June 2-27, 2014*, 2014, [arXiv:1411.4085].

[21] S. Catani et al., "QCD matrix elements + parton showers", *JHEP* **11** (2001) p. 063, DOI: 10.1088/1126-6708/2001/11/063, [arXiv:hep-ph/0109231].

[22] M. L. Mangano, M. Moretti and R. Pittau, "Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b}$ + n jets as a case study", *Nucl. Phys.* **B632** (2002) 343–362, DOI: 10.1016/S0550-3213(02)00249-3, [arXiv:hep-ph/0108069].

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

[24] UA1 Collaboration, "Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider", *Phys. Lett.* **B126** (1983) 398–410, DOI: 10.1016/0370-2693(83)90188-0.

[25] UA2 Collaboration, "Evidence for $Z^0 \to e^-e^+$ at the CERN $\bar{p}p$ Collider", *Phys. Lett.* **B129** (1983) 130–140, DOI: 10.1016/0370-2693(83)90744-X.

[26] LEP Electroweak Working Group, "Precision Electroweak Measurements and Constraints on the Standard Model" (2010), [arXiv:1012.2367].

[27] S. D. Drell and T.-M. Yan, "Massive Lepton Pair Production in Hadron-Hadron Collisions at High-Energies", *Phys. Rev. Lett.* **25** (1970) 316–320, DOI: 10.1103/PhysRevLett.25.316.

[28] L. Evans and P. Bryant, "LHC Machine", *Journal of Instrumentation* **3** (2008) S08001, DOI: 10.1088/1748-0221/3/08/S08001.

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

[30] 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) 1–29, DOI: 10.1016/j.physletb.2012.08.020, [arXiv:1207.7214].

[31] CERN, *LEP Design Report: Vol.2. The LEP Main Ring*, 1984, CERN-LEP-84-01.

[32] C. Lefèvre, "The CERN accelerator complex. Complexe des accélérateurs du CERN", 2008, URL: http://cds.cern.ch/record/1260465/.

[33] CMS Collaboration, "The CMS experiment at the CERN LHC", *Journal of Instrumentation* **3 S08004** (2008) p. 361, DOI: 10.1088/1748-0221/3/08/S08004.

[34] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008) S08003, DOI: 10.1088/1748-0221/3/08/S08003.

[35] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *Journal of Instrumentation* **3** (2008) S08002, DOI: 10.1088/1748-0221/3/08/S08002.

[36] LHCb Collaboration, "The LHCb Detector at the LHC", *Journal of Instrumentation* **3** (2008) S08005, DOI: 10.1088/1748-0221/3/08/S08005.

[37] TOTEM Collaboration, "The TOTEM Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008) S08007, DOI: 10.1088/1748-0221/3/08/S08007.

[38] LHCf Collaboration, "The LHCf detector at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008) S08006, DOI: 10.1088/1748-0221/3/08/S08006.

[39] M Bajko et al., *Report of the Task Force on the Incident of 19th September 2008 at the LHC*, tech. rep., Geneva: CERN, 2009, LHC-PROJECT-Report-1168.

[40] CMS Collaboration, *Public CMS Luminosity Information - Run 1 Annual Charts of Luminosity*, URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#Run_1_Annual_Charts_of_Luminosit (visited on 05/01/2016).

[41] The HiLumi LHC Collaboration, "HL-LHC Preliminary Design Report: Deliverable: D1.5" (2014), CERN-ACC-2014-0300.

[42] CMS Collaboration, *CMS People Statistics*, URL: http://cms.web.cern.ch/content/people-statistics (visited on 05/04/2016).

[43] J. Berger, "Search for the Higgs Boson Produced via Vector-Boson Fusion in the Decay Channel H $\to \tau\tau$", PhD thesis, KIT, 2014, IEKP-KA/2014-10.

[44] CMS Collaboration, *3D models of the CMS detector and events in SketchUp*, URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/SketchUpCMS (visited on 10/02/2016).

[45] F. M. Stober, "Measurement of the Three-Jet Mass Cross-Section at $\sqrt{s} = 7\,\text{TeV}$", PhD thesis, KIT, 2012, IEKP-KA/2012-24.

[46] CMS Collaboration, "Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in *pp* Collisions at $\sqrt{s} = 7$ TeV", *JINST* **8** (2013) P09009, DOI: 10.1088/1748-0221/8/09/P09009, [arXiv:1306.2016].

[47] V. D. Elvira, *Measurement of the Pion Energy Response and Resolution in the CMS HCAL Test Beam 2002 Experiment*, CMS Note, Geneva: CERN, 2004, CMS-NOTE-2004-020.

[48] CMS Collaboration, "The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7\,\text{TeV}$ at the LHC", *JINST* **8** (2013) P11002, DOI: 10.1088/1748-0221/8/11/P11002, [arXiv:1306.6905].

[49] CMS Collaboration, "Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV", CMS Physics Analyses Summary (2010), CMS-PAS-PFT-10-002.

[50] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS Physics Analyses Summary (2009), CMS-PAS-PFT-09-001.

[51] F. Beaudette, "The CMS Particle Flow Algorithm", *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013)*, 2013, 295–304, [arXiv:1401.8155].

[52] R. Frühwirth, "Application of Kalman filtering to track and vertex fitting", *Nucl. Instrum. Meth.* **A262** (1987) 444–450, DOI: 10.1016/0168-9002(87)90887-4.

[53] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *Transactions of the ASME–Journal of Basic Engineering* **82** (1960) 35–45.

[54] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014) P10009, DOI: 10.1088/1748-0221/9/10/P10009, [arXiv:1405.6569].

[55] CMS Collaboration, "Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, DOI: 10.1088/1748-0221/7/10/P10002, [arXiv:1206.4071].

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

[57] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8\,\text{TeV}$", *JINST* **10** (2015) P06005, DOI: 10.1088/1748-0221/10/06/P06005, [arXiv:1502.02701].

[58] W. Adam et al., *Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC*, CMS Note, Geneva: CERN, 2005, CMS-NOTE-2005-001.

[59] J. Berger, "Analysis of $Z \to \mu\mu$ + Jet Events at the LHC", diploma thesis, KIT, 2010, IEKP-KA/2010-23.

[60] G. P. Salam and G. Soyez, "A practical seedless infrared-safe cone jet algorithm", *JHEP* **05** (2007) p. 086 [arXiv:0704.0292].

[61] S. Catani, Y. L. Dokshitzer and B. R. Webber, "The $K_\perp$ clustering algorithm for jets in deep inelastic scattering and hadron collisions", *Phys. Lett.* **B285** (1992) 291–299, DOI: 10.1016/0370-2693(92)91467-N.

[62] Y. L. Dokshitzer et al., "Better jet clustering algorithms", *JHEP* **08** (1997) p. 001, DOI: 10.1088/1126-6708/1997/08/001, [arXiv:hep-ph/9707323].

[63] M. Wobisch and T. Wengler, "Hadronization corrections to jet cross-sections in deep inelastic scattering", *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, 1998 [arXiv:hep-ph/9907280].

[64] M. Cacciari, G. P. Salam and G. Soyez, "The anti-k_t jet clustering algorithm", *Journal of High Energy Physics* (2008) p. 063 [arXiv:0802.1189].

[65] ATLAS Collaboration, "Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7\,\text{TeV}$", *Eur. Phys. J.* **C73** (2013) p. 2304, DOI: 10.1140/epjc/s10052-013-2304-2 [arXiv:1112.6426].

[66] K. Rabbertz, personal communication, 2016, URL: https://hypernews.cern.ch/HyperNews/CMS/get/JME-13-004/49.html.

[67] M. Cacciari, G. P. Salam and G. Soyez, "The Catchment Area of Jets", *JHEP* **04** (2008) p. 005, DOI: 10.1088/1126-6708/2008/04/005 [arXiv:0802.1188].

[68] CMS Collaboration, "Measurement of the underlying event activity in *pp* collisions at $\sqrt{s} = 0.9$ and 7 TeV with the novel jet-area/median approach", *JHEP* **08** (2012) p. 130, DOI: 10.1007/JHEP08(2012)130 [arXiv:1207.2392].

[69] CMS Collaboration, "Missing transverse energy performance of the CMS detector", *JINST* **6** (2011) P09001, DOI: 10.1088/1748-0221/6/09/P09001, [arXiv:1106.5048].

[70] CMS Collaboration, "MET performance in 8 TeV data", CMS Physics Analyses Summary (2013), CMS-PAS-JME-12-002.

[71] CMS Collaboration, "Pileup Jet Identification", CMS Physics Analyses Summary (2013), CMS-PAS-JME-13-005.

[72] D. Bertolini et al., "Pileup Per Particle Identification", *JHEP* **10** (2014) p. 59, DOI: 10.1007/JHEP10(2014)059, [arXiv:1407.6013].

[73] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update", CMS Physics Analyses Summary (2013), CMS-PAS-LUM-13-001.

[74] S. van der Meer, *Calibration of the effective beam height in the ISR.* preprint, Geneva: CERN, 1968, ISR-PO-68-31.

[75] Worldwide LHC Computing Grid, http://wlcg-public.web.cern.ch.

[76] ROOT — A Data Analysis Framework, http://root.cern.ch/.

[77] R. Brun and F. Rademakers, "ROOT - An Object Oriented Data Analysis Framework", *Nucl. Instrum. Meth. A* **389** (1997) 81–86, DOI: 10.1016/S0168-9002(97)00048-X.

[78] CMSSW Application Framework, https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework.

[79] Kappa - Karlsruhe Advanced Package for Physics Analysis, https://github.com/KappaAnalysis/Kappa.

[80] J. Berger et al., "ARTUS - A Framework for Event-based Data Analysis in High Energy Physics" (2015), [arXiv:1511.00852].

[81] Excalibur - Analysis repository for Z+Jet studies, https://github.com/artus-analysis/Excalibur.

[82] G. Sieber, "Measurement of Triple-Differential Dijet Cross Sections with the CMS Detector at 8 TeV and PDF Constraints", in preparation, PhD thesis, KIT, 2016.

[83] T. Müller, "Analysis of Standard Model Higgs Boson Decays to Tau Pairs with the CMS Detector at the LHC", PhD thesis, KIT, Karlsruhe, 2015, IEKP-KA/2015-24.

[84] Grid-Control, https://ekptrac.physik.uni-karlsruhe.de/trac/grid-control/.

[85] C. Metzlaff, "Optimierung eines Cachingsystems für einen hohen Datendurchsatz bei der Analyse von *pp*-Kollisionen am LHC", MA thesis, KIT, 2016, IEKP-KA/2016-02.

[86] M. Fischer et al., "Tier 3 batch system data locality via managed caches", *Proceedings, 16th International workshop on Advanced Computing and Analysis Techniques in physics (ACAT 14)*, vol. 608, 2015, p. 012018, DOI: 10.1088/1742-6596/608/1/012018.

[87] YODA — Yet more Objects for Data Analysis, https://yoda.hepforge.org/.

[88] matplotlib: python plotting — Matplotlib 1.2.0 documentation, http://matplotlib.org/.

[89] A. Buckley et al., "General-purpose event generators for LHC physics", *Phys. Rept.* **504** (2011) 145–233, DOI: 10.1016/j.physrep.2011.03.005, [arXiv:1101.2599].

[90] T. Gleisberg et al., "Event generation with SHERPA 1.1", *JHEP* **02** (2009) p. 007, DOI: 10.1088/1126-6708/2009/02/007, [arXiv:0811.4622].

[91] T. Sjöstrand et al., "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159–177, DOI: 10.1016/j.cpc.2015.01.024, [arXiv:1410.3012].

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

[93] J. Alwall et al., "MadGraph 5: going beyond", *JHEP* **06** (2011) 128, DOI: 10.1007/JHEP06(2011)128 [arXiv:1106.0522v1].

[94] R. Gavin et al., "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order", *Comput. Phys. Commun.* **182** (2011) 2388–2403, DOI: 10.1016/j.cpc.2011.06.008 [arXiv:1011.3540].

[95] S. Agostinelli et al., "Geant4—a simulation toolkit", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003) 250 –303, DOI: 10.1016/S0168-9002(03)01368-8.

[96] G. Fleig, "Dynamic Integration of Cloud Resources into Local Computing Clusters and Calibration of the Jet Energy Scale with the CMS Detector at $\sqrt{s} = 13\,\text{TeV}$", MA thesis, KIT, 2016, IEKP-KA/2016-03.

[97] SheRivF toolkit, https://github.com/dhaitz/SheRivF.

[98] T. Kluge, K. Rabbertz and M. Wobisch, "FastNLO: Fast pQCD calculations for PDF fits", *Deep inelastic scattering. Proceedings, 14th International Workshop, DIS 2006, Tsukuba, Japan, April 20-24, 2006*, 2006, 483–486, [arXiv:hep-ph/0609285].

[99] A. Buckley et al., "LHAPDF6: 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].

[100] L. Del Debbio, N. P. Hartland and S. Schumann, "MCgrid: projecting cross section calculations on grids", *Comput. Phys. Commun.* **185** (2014) 2115–2126, DOI: 10.1016/j.cpc.2014.03.023, [arXiv:1312.4460].

[101] E. Bothmann, M. Schönherr and S. Schumann, "Fast evaluation of theoretical uncertainties with Sherpa and MCgrid", *23rd International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2015) Dallas, Texas, United States, April 27-May 1, 2015*, vol. DIS2015, 2015, p. 136, [arXiv:1507.03908].

[102] S. Alekhin et al., "HERAFitter", *Eur. Phys. J.* **C75** (2015) p. 304, DOI: 10.1140/epjc/s10052-015-3480-z, [arXiv:1410.4412].

[103] xFitter, https://wiki-zeuthen.desy.de/xFitter/.

[104] S. Fanchiotti, B. Kniehl and A. Sirlin, "Incorporation of QCD effects in basic corrections of the electroweak theory", *Phys. Rev. D* **48** (1993) 307–331, DOI: 10.1103/PhysRevD.48.307 [arXiv:hep-ph/9212285].

[105] M. Botje, "QCDNUM: Fast QCD Evolution and Convolution", *Comput. Phys. Commun.* **182** (2011) 490–532, DOI: 10.1016/j.cpc.2010.10.020, [arXiv:1005.1481].

[106] F. James and M. Roos, "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations", *Comput. Phys. Commun.* **10** (1975) 343–367, DOI: 10.1016/0010-4655(75)90039-9.

[107] CMS Collaboration, "Measurement of the associated production of a Z boson and b quarks in pp collision at 8 TeV", CMS Physics Analyses Summary (2015), CMS-PAS-SMP-14-010.

[108] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *Journal of Instrumentation* **6** (2011) P11002, DOI: 10.1088/1748-0221/6/11/P11002 [arXiv:1107.4277].

[109] CMS Collaboration, "Jet energy calibration in the 8 TeV *pp* data", *Journal of Instrumentation* (2016), to be published.

[110] CMS Collaboration, *Jet energy scale performance in 2011*, CMS Detector Performance Summary, 2012, CMS-DP-2012-006.

[111] CMS Collaboration, *Jet energy corrections and uncertainties. Detector performance plots for 2012*, CMS Detector Performance Summary, 2012, CMS-DP-2012-012.

[112] CMS Collaboration, *Status of the 8 TeV jet energy corrections and uncertainties based on 11 fb^{-1} of data in CMS*, CMS Detector Performance Summary, 2013, CMS-DP-2013-011.

[113] CMS Collaboration, *8 TeV jet energy corrections and uncertainties based on 19.8 fb^{-1} of data in CMS*, CMS Detector Performance Summary, 2013, CMS-DP-2013-033.

[114] CMS Collaboration, *CMS JEC Run I legacy performance plots*, CMS Detector Performance Summary, 2015, CMS-DP-2012-012.

[115] D. Haitz, "Measurement of Jet Energy Scale and Resolution at ATLAS and CMS at $\sqrt{s} = 8\,\text{TeV}$ TeV", *Proceedings, 16th conference on Elastic and Diffractive Scattering (EDS 15)*, vol. 8, 2015, p. 941, DOI: 10.5506/APhysPolBSupp.8.941.

[116] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Physics Letters B* **659** (2008) 119 –126, DOI: 10.1016/j.physletb.2007.09.077 [arXiv:0707.1378].

[117] M. Fischer, "Coordinated Caching for High Performance Calibration using $Z \to \mu\mu$ Events of the CMS Experiment", in preparation, PhD thesis, 2016.

[118] J. Handl, "Jet Energy Calibration with $Z(\to e^+e^-)$+jet events at $\sqrt{s} = 13\,\text{TeV}$ ", in preparation, Bachelor's thesis, 2016.

[119] D. Haitz, "Analyse von $Z(\to \mu^+\mu^-)$+Jet Ereignissen und Kalibration der Jet-Energieskala des CMS-Detektors", diploma thesis, KIT, 2013, IEKP-KA/2013-18.

[120] T. Hauth, "Dynamic Extension of Local Batch Systems with Cloud Resources and Measurement of the Jet Energy Scale with the CMS Detector", diploma thesis, KIT, 2011, IEKP-KA/2011-7.

[121] D. Piparo, "Statistical Combination of Higgs Decay Channels and Determination of the Jet-Energy Scale of the CMS Experiment at the LHC", IEKP-KA/2010-22, PhD thesis, KIT, IEKP-KA/2010-22.

[122] V. Büge, "Virtualisation of Grid Resources and Prospects of the Measurement of Z Boson Production in Association with Jets at the LHC", IEKP-KA/2008-18, PhD thesis, KIT, IEKP-KA/2008-18.

[123] S. Torbjörn S. Mrenna and P. Skands, "PYTHIA 6.4 physics and manual", *Journal of High Energy Physics* (2006) p. 026, DOI: 10.1088/1126-6708/2006/05/026 [arXiv:hep-ph/0603175].

[124] CMS Collaboration, *Jet Flavour Identification*, URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideBTagMCTools (visited on 05/04/2016).

[125] CMS Collaboration, "Performance of quark/gluon discrimination in 8 TeV pp data", CMS Physics Analyses Summary (2013), CMS-PAS-JME-13-002.

[126] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, DOI: 10.1088/1748-0221/8/04/P04013, [arXiv:1211.4462v2].

[127] F. D. Aaron et al., "Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA", *JHEP* **01** (2010) p. 109, DOI: 10.1007/JHEP01(2010)109, [arXiv:0911.0884].

[128] D. Haitz, "PDF Constraints from $Z(\rightarrow e^+e^-)$+Jet Double-Differential Cross Section Measurement in p_T and $|y|$ at $\sqrt{s} = 8\,\text{TeV}$" (2016), CMS-AN-2016-114.

[129] C. F. Berger et al., "An Automated Implementation of On-Shell Methods for One-Loop Amplitudes", *Phys. Rev.* **D78** (2008) p. 036003, DOI: 10.1103/PhysRevD.78.036003, [arXiv:0803.4180].

[130] A. Buckley et al., "Rivet user manual", *Comput. Phys. Commun.* **184** (2013) 2803–2819, DOI: 10.1016/j.cpc.2013.05.021, [arXiv:1003.0694].

[131] L. K. Saini, I. Kravchenko and Y. Maravin, *A study of efficiencies and scale factors for cut-based electron identification at CMS experiment using data from proton-proton collisions at* $\sqrt{s} = 8\,\text{TeV}$. 2014, CMS-AN-2014-055.

[132] J. M. Campbell and R. K. Ellis, "An Update on vector boson pair production at hadron colliders", *Phys. Rev.* **D60** (1999) p. 113006, DOI: 10.1103/PhysRevD.60.113006, [arXiv:hep-ph/9905386].

[133] M. Czakon and A. Mitov, "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders", *Comput. Phys. Commun.* **185** (2014) p. 2930, DOI: 10.1016/j.cpc.2014.06.021, [arXiv:1112.5675].

[134] N. Kidonakis, "Top Quark Production", *Proceedings, Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013)*, 2014, 139–168, DOI: 10.3204/DESY-PROC-2013-03/Kidonakis, [arXiv:1311.0283].

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

[136] T. 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, 6 p [arXiv:1105.1160].

[137] S. Alioli et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) p. 043, DOI: 10.1007/JHEP06(2010)043 [arXiv:1002.2581].

[138] CMS Collaboration, "Measurement of the Z boson differential cross section in transverse momentum and rapidity in proton–proton collisions at 8 TeV", *Phys. Lett.* **B749** (2015) 187–209, DOI: 10.1016/j.physletb.2015.07.065, [arXiv:1504.03511].

[139] R. Boughezal, X. Liu and F. Petriello, "Phenomenology of the Z-boson plus jet process at NNLO" (2016), [arXiv:1602.08140].

[140] ATLAS Collaboration, "Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector", *JHEP* **09** (2014) p. 145, DOI: 10.1007/JHEP09(2014)145, [arXiv:1406.3660].

[141] W. Hollik et al., "Electroweak corrections to Z-boson hadroproduction at finite transverse momentum", *Nucl. Phys.* **B900** (2015) 576–602, DOI: 10.1016/j.nuclphysb.2015.09.018, [arXiv:1504.07574].

[142] G. Salam, *LHC Physics at 1% precision?*, Seminar at SLAC, 2016, URL: https://gsalam.web.cern.ch/gsalam/talks/repo/2016-03-SB+SLAC-SLAC-precision.pdf.

[143] CMS Collaboration, "Measurement of the double-differential inclusive jet cross section at $\sqrt{s} = 8\,\text{TeV}$", CMS Physics Analyses Summary (2015), CMS-PAS-SMP-14-001.

[144] R. S. Thorne, "Effect of changes of variable flavor number scheme on parton distribution functions and predicted cross sections", *Phys. Rev.* **D86** (2012) p. 074017, DOI: 10.1103/PhysRevD.86.074017, [arXiv:1201.6180].

[145] 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].

[146] S. Camarda et al., "QCD analysis of W- and Z-boson production at Tevatron", *Eur. Phys. J.* **C75** (2015) p. 458, DOI: 10.1140/epjc/s10052-015-3655-7, [arXiv:1503.05221].

[147] ATLAS Collaboration, "Measurement of the inclusive jet cross section in pp collisions at $\sqrt{s} = 2.76\,\text{TeV}$ and comparison to the inclusive jet cross section at $\sqrt{s} = 7\,\text{TeV}$ using the ATLAS detector", *Eur. Phys. J.* **C73** (2013) p. 2509, DOI: 10.1140/epjc/s10052-013-2509-4, [arXiv:1304.4739].

[148] CMS Collaboration, *Jet Identification*, URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID (visited on 05/04/2016).

[149] CMS Collaboration, Baseline muon selections for Run-I, URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId (visited on 05/04/2016).

[150] CMS Collaboration, *Electron ID Working Points*, URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/EgammaCutBasedIdentification_Electron_ID_Working_Points (visited on 16/02/2016).

Danksagung

Der höchste Dank an dieser Stelle gebührt Prof. Dr. Günter Quast: für die individuelle Betreuung und Förderung. die fachliche und organisatorische Unterstützung, das mir entgegengebrachte Vertrauen und das Leiten der Arbeitsgruppe mit Begeisterung, Einsatz, Weitblick und Menschlichkeit. Dankbar blicke ich auf vier spannende und lehrreiche Jahre zurück.

Prof. Dr. Wim de Boer danke ich für die freundliche Übernahme des Korreferats und die hilfreichen Kommentare bei dieser Dissertation sowie bereits bei meiner Diplomarbeit.

Dr. Klaus Rabbertz danke ich vielmals für die Betreuung und Unterstützung bei technischen und physikalischen Fragestellungen, die maßgeblich zum Gelingen dieser Arbeit beigetragen haben. Vielen Dank auch an Georg Sieber, mit dem ich unzählige konstruktive Diskussionen führte und dessen Wissen über QCD und PDFs ich immer anzapfen durfte.

Bei Dr. Joram Berger möchte ich mich herzlichst für die jahrelange harmonische sowie produktive Betreuung und Zusammenarbeit bei der Jetkalibration bedanken. Für viele weitere Anregungen und Diskussionen bedanke ich mich bei Mikko Voutilainen.

Ein zusätzlicher Dank geht an die übrigen Mitglieder der Arbeitsgruppe und des Instituts, insbesondere den Korrekturlesern und den Admins für ihren Einsatz.

Meiner Freundin Mareike sowie meinen Eltern Paul und Monika gebührt ein ganz besonderer Dank, für das Vertrauen und die Unterstützung weit über diese Arbeit hinaus.

Hiermit versichere ich, die vorliegende Arbeit selbstständig verfasst und nur die angegebenen Hilfsmittel verwendet zu haben.

Dominik Haitz

Karlsruhe, den 04.05.2016