Karlsruher Institut für Technologie

IEKP-KA/2019-16

Jet energy calibration and triple differential inclusive cross section measurements with Z ($\rightarrow \mu\mu$) + jet events at 13 TeV recorded by the CMS detector

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN (Dr. rer. nat.)

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

DISSERTATION

von

Dipl.-Phys. Thomas Berger
aus Saarlouis

Tag der mündlichen Prüfung: 25. 10. 2019

Referent: Prof. Dr. Günter Quast
Korreferent: Priv.-Doz. Dr. Klaus Rabbertz

Abstract

This thesis presents two substantial aspects of the analysis of Z boson production in association with jets, where the Z boson decays to a muon-antimuon pair: jet energy calibration and triple differential cross section measurements. Data have been used that were recorded with the CMS detector in the year 2016 at a center-of-mass energy of 13 TeV. These data correspond to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$ of proton-proton collisions at the LHC.

The first part of this thesis contributes to the jet energy calibration in CMS by estimating residual corrections that account for differences in jet reconstruction between measurement and simulation. Correction factors are extracted from Z ($\rightarrow \mu\mu$) + jet events by comparing the transverse momentum of a jet to the one of a balancing Z boson, which can be reconstructed very precisely from muons.

The second part of this thesis presents triple differential measurements of inclusive cross sections for Z ($\rightarrow \mu\mu$) + jet production. The cross sections are measured as a function of the Z boson transverse momentum $p_{\mathrm{T}}^{\mathrm{Z}}$ or alternatively the variable ϕ_{η}^{*}, the rapidity separation y^{*} of the Z boson and the leading jet, and the boost y_{b} of their center-of-mass system. The observable ϕ_{η}^{*} has the advantage to be determined from angular information of the muons only, thereby further increasing the reconstruction precision in comparison to $p_{\mathrm{T}}^{\mathrm{Z}}$. The observables y^{*} and y_{b} introduce a division of the phase space that promises a better sensitivity to the parton subprocesses and thus the PDFs. The measured cross sections are corrected for detector effects by a three-dimensional unfolding procedure and are compared to theory predictions calculated at next-to-next-to-leading order in perturbative QCD.

Contents

"Daß ich erkenne, was die Welt
Im Innersten zusammenhält"

Johann Wolfgang von Goethe,
Faust I

CHAPTER 1

Introduction

"So that I may perceive whatever holds
The world together in its inmost folds"

The above quote from Johann Wolfgang von Goethe's *Faust* describes the fascination and enthusiasm that drives many researchers. But instead of making a pact with the devil, as the protagonist of the play does, modern science aims to develop technological devices and theoretical concepts that help improve our understanding of the world.

In physics, this has led to the construction of the Large Hadron Collider (LHC) at the CERN research centre near Geneva, the most powerful and advanced particle accelerator built to day, which is used to test one of the most sophisticated theoretical frameworks, the Standard Model of particle physics. An important missing piece of the Standard Model, a Higgs boson, whose existence was predicted almost 50 years earlier, was discovered at the LHC in the year 2012, which lead to the award of the Nobel Prize in the year 2013 [1].

However, particle physics still holds secrets to be unveiled. One of them is the inner structure of the proton. The Standard Model provides techniques for calculating processes involving elementary particles. However, the description of processes involving composite particles such as the proton requires knowledge of how their constituents, the partons, contribute to the overall process. This information can be provided by parton distribution functions (PDFs).

The PDFs are a measure of the probability of finding a parton with a certain fraction of the momentum of the proton. The lack of a sophisticated technique which permits to compute PDFs from first principles necessitates the use of experimental data to determine them. Latest results have been obtained by combining data from various scattering processes such as electron-proton scattering at the HERA collider [2] at the DESY research centre in Hamburg, proton-antiproton scattering at the Tevatron collider at the Fermilab research centre and proton-proton scattering at the LHC [3–6]. Particle physics, however, has reached a point at which measurements have become more precise than the calculations of certain processes. Further constraints on PDFs which reduce their uncertainties can therefore help to improve predictions for many physics analyses.

High experimental precision is required for this task. The CMS detector at the LHC achieves its highest precision in the reconstruction of muons and antimuons. Processes in which Z bosons decay into muon-antimuon pairs can therefore be used for two main tasks on which this thesis is focused. First of all, the measurement of cross sections for Z boson production can be used for deriving PDF constraints. Secondly, the analysis of Z boson events also has a technical application in the calibration of jets, the observable signature of partons in the detector.

This thesis presents both aspects of a Z boson analysis. After introducing the necessary theoretical background and explaining the experimental setup, the CMS detector, the first part discusses the steps of the jet energy calibration of the data taken in the year 2016. The work of this thesis contributed to the estimation of residual corrections, which account for differences between measurement and simulation.

The second part presents the measurement of the triple differential inclusive Z boson cross section in association with one jet. The triple differential measurement uses observables which allow for a suitable division of the phase space in order to obtain a better sensitivity to the parton subprocesses and thus the PDFs. To correct the measured cross sections for detector effects, a three-dimensional unfolding procedure is used. The comparison of the resulting cross sections to theory predictions calculated at next-to-next-to-leading order perturbative QCD provides information on the constraints on the PDFs that can be concluded from this measurement.

CHAPTER 2

Theoretical background

To introduce the necessary theoretical background for this thesis, the following chapter presents the basic ideas of the physical and mathematical structure on which particle physics is based on. In a first step, a general introduction into the phenomenology of the Standard Model is given in Section 2.1, followed by the basic ideas of quantum field theory in Section 2.2. Subsequently, the main aspects of both the electroweak theory and quantum chromodynamics are discussed in Sections 2.3 and 2.4, respectively. The methods of Monte Carlo simulations are explained in Section 2.5. Finally, the techniques of jet clustering are introduced in Section 2.6.

2.1 Phenomenology of the Standard Model of particle physics

The discovery of the neutron completed a very simple concept of particle physics in the first half of the 20th century, according to which matter is composed only of electrons, protons and neutrons. Atoms consist of electrons and an atomic nucleus, which itself consists of protons and neutrons. However, this model could not be sustained when new particles were discovered subsequently, such as the muon or the neutrino. New concepts have been developed, such as the quark model, which can be used to understand protons and neutrons as composite particles. These concepts are nowadays unified in the *Standard Model* of particle physics, based on the small set of elementary particles presented in Figure 2.1, mathematically embedded in the structure of quantum field theory.

There are two classes of particles: *fermions*, the building blocks of matter, and *bosons*, the mediators of interactions. There are three fundamental interactions which can be described by the Standard Model: the *electromagnetic* force, the *weak* force and the *strong* force. Each elementary boson corresponds to one of the forces, each elementary fermion can be classified via the forces acting on it. Figure 2.1 gives an overview of the elementary particles of the Standard Model, whose properties are introduced in the following.

Fermions that are subject to the strong force are called *quarks*, the bosons that mediate the strong force are called *gluons*. Most matter in the universe consists of *up* and *down* quarks, though it has been found that each quark has two equivalents with the same properties except their masses, also referred to as different quark *flavours*. The strong

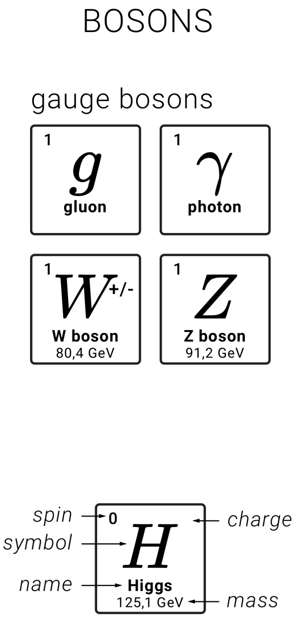


Figure 2.1: Elementary particles that occur in the Standard Model of particle physics. Illustration provided by [7] based on properties taken from [8].

force holds *protons* and *neutrons* together and therefore plays an important role in nuclear reactions. Particles that are subject to the strong force carry *colour* charge. The theory that describes the strong force is called quantum chromodynamics and will be discussed in Section 2.4.

Fermions that are not subject to the strong force are called *leptons*, whose most famous representative is the *electron*. Similar to quarks, each lepton has two equivalents sharing the same properties except for their masses. The *muon* should be mentioned as the heavier equivalent of the electron because of its particular importance to this thesis.

Particles experiencing the electromagnetic force carry electromagnetic charge, whose mediating boson is the *photon*. The electromagnetic force defines the properties of atoms and molecules and is therefore the most abundant fundamental force in everyday life (besides *gravity*, which is not part of the Standard Model).

Leptons that are not subject to the electromagnetic interaction are called *neutrinos*. Within the Standard Model, neutrino masses are assumed to be zero. Although this assuption is known to violate the observation of neutrino oscillations, it is a suitable approximation given the limit provided by the Particle Data Group Collaboration [8] in comparison to other particle masses displayed in Figure 2.1.

All fermions experience the weak force, mediated by the Z and W bosons. It is the only interaction that provides the possibility to change the flavour of a particle and therefore plays a crucial role in the decays of atomic nuclei. The electromagnetic and the weak interaction have been unified within the *electroweak* interaction, which will be discussed in Section 2.3.

Before its discovery in 2012, the *Higgs* boson was the last missing piece to complete the Standard Model [9]. A particle acquires mass by interacting with the Higgs field, a mechanism that has already been published in the 1960s [10].

2.2 Quantum field theory

Quantum field theory (QFT) provides the mathematical structure of the Standard Model. It combines quantum mechanics and field theory by considering quantum mechanical operators as functions of continuous parameters such as space-time. Field theories rely on a *Lagrangian* $\mathcal{L}$ that defines the equations of motion derived as *Euler-Lagrange* equations.

For a fermion of mass m, described by a spinor ψ, the Lagrangian $\mathcal{L}_\mathrm{f}$ yields the *Dirac* equation.

$$\mathcal{L}_\mathrm{f} = \overline{\psi}(\mathrm{i}\not{\partial} - m)\psi \tag{2.1}$$

To ensure *gauge* symmetry, which is one of the fundamental principles of theoretical particle physics and well known for classical electrodynamics, the Lagrangian has to be modified by the introduction of gauge fields $\mathcal{A}_\mu$ with vector boson properties. The Standard Model Lagrangian is invariant under transformations belonging to the symmetry groups U(1), SU(2) and SU(3), which are connected to the fundamental forces as summarized in Table 2.1. The vector bosons can be related to the generators of the gauge group and fulfil the Maxwell equations derived from the Lagrangian $\mathcal{L}_\mathrm{b}$, written with the field tensor $\mathcal{F}_{\mu\nu} = \partial_\mu \mathcal{A}_\nu - \partial_\nu \mathcal{A}_\mu$.

$$\mathcal{L}_\mathrm{b} = -\frac{1}{4}\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu} \tag{2.2}$$

The full Lagrangian follows as the sum of $\mathcal{L}_\mathrm{f}$ and $\mathcal{L}_\mathrm{b}$ by replacing all derivatives ∂_μ with the covariant derivatives $D_\mu = \partial_\mu - \mathrm{i}g\mathcal{A}_\mu$. This procedure is known as *minimal coupling*. It creates interaction terms containing the coupling constant g in addition to the free field Lagrangians $\mathcal{L}_\mathrm{f}$ and $\mathcal{L}_\mathrm{b}$. In an abelian gauge theory, the interaction Lagrangian is given by $\mathcal{L}_\mathrm{int}$, although additional terms can follow in non-abelian gauge theories when replacing ∂_μ by D_μ in the field tensor $\mathcal{F}_{\mu\nu}$ (*Yang-Mills-Theory*).

$$\mathcal{L}_\mathrm{int} = g\overline{\psi}\not{\mathcal{A}}\psi \tag{2.3}$$

In particle physics, the common experimental setup is scattering, meaning that a set of incoming particles interacts to form a set of outgoing particles, whose properties can then be estimated by the measurement.

Following from the quantum nature of QFT, a theoretical prediction determines the probability that a particular configuration of the properties of outgoing particles is formed. This probability divided by the flux of incoming particles is defined as the *cross section*, which is the central observable in both experimental and theoretical particle physics. The

Table 2.1: Gauge symmetries and the fundamental forces.

gauge group	number of generators	fundamental force	gauge boson
U(1)	1	electromagnetic force	photon
SU(2)	3	weak force	W^+, W^-, Z boson
SU(3)	8	strong force	8 gluons

flux of particles, more specific the number of particles passing through an area per time, is also known as *luminosity*.

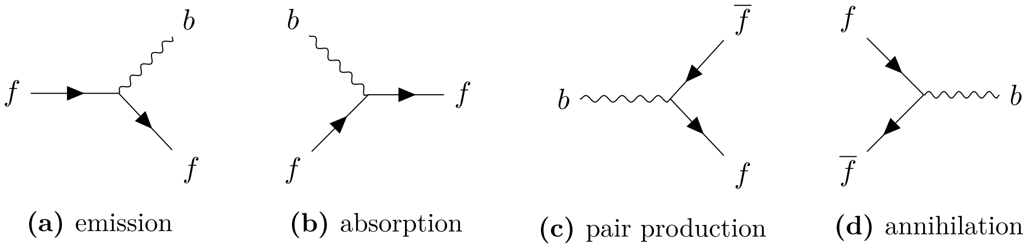
In scattering theory, the incoming and the outgoing particles are assumed to behave asymptotically like free, non-interacting particles and therefore follow the free Lagrangian $\mathcal{L}_\mathrm{f} + \mathcal{L}_\mathrm{b}$, whose equations of motion can be solved analytically. If the coupling constant is small, the impact of the interaction terms on the time evolution can be treated perturbatively. The terms contributing to the calculation of matrix elements up to a particular order in perturbation theory can be visualized in *Feynman diagrams*, and reversely, the diagrams can be interpreted as mathematical expressions using *Feynman rules*. Feynman rules are derived from the structure of the Lagrangian. The two basic concepts are the propagation of particles illustrated as lines, following from the free Lagrangian, and their interactions illustrated by vertices, following from the interaction Lagrangian. Figure 2.2 shows the four basic processes that follow from $\mathcal{L}_\mathrm{int}$ in Equation 2.3. The full set of Feynman rules can be found in textbooks about QFT such as [11].

Feynman diagrams are classified as *leading order* (LO), *next-to-leading order* (NLO), *next-to-next-to-leading order* (NNLO), etc., according to the power of the coupling constant g in the perturbative series. Since each vertex corresponds to a factor of g, the number of vertices represents the order at which a certain diagram contributes to the process.

At LO, the lowest number of vertices is required to visualize a process of interest in the simplest way. Feynman diagrams contributing to calculations at LO typically show a tree-like structure. At higher orders, calculations become challenging for diagrams in which additional vertices allow the formation of loops. Loops are interpreted as integrals which are usually divergent.

At NLO, there are divergent terms from Feynman diagrams containing physically non-observable emissions of bosons which are *soft* (carrying little energy) or *collinear* (emitted at narrow angles). These terms exhibit the same functional structure as the low-energy region of the loop integration phase space (*infrared divergence*). The corresponding diagrams are referred to as *real* and *virtual* corrections to the LO calculation, carry different signs and therefore cancel each other out.

However, divergencies at the high-energy region (ultraviolet divergence) are absorbed into free parameters such as the coupling constant, a procedure called *renormalization*. This comes at the price of the introduction of an arbitrarily chosen renormalization scale μ_R, on which the renormalized parameters depend. Measurable physical quantities are supposed not to depend on an arbitrarily chosen parameter, which is only an artifact of the mathematical approach used for the calculation. In fact, the scale dependence



(a) emission **(b)** absorption **(c)** pair production **(d)** annihilation

Figure 2.2: Basic processes described by the interaction Lagrangian $\mathcal{L}_{\text{int}}$ from Equation 2.3. It defines the vertex as the crossing point of two fermion lines f and one boson line b. If the horizontal is understood as the time axis, four processes can be illustrated by the same vertex: **(a)** a fermion emits a boson; **(b)** a fermion absorbs a boson; **(c)** a boson splits into a fermion and an antifermion; **(d)** a fermion and an antifermion fuse into a boson.

decreases when including higher orders in perturbation theory. It is expected to eventually vanish completely, but remains for calculations at fixed orders in the form of a scale uncertainty.

The ratio between results derived at two consecutive orders (for example NLO/LO or NNLO/NLO) is referred to as (NLO or NNLO) k-factor. A perturbative series can be checked for convergence by assuring that the k-factors converge to unity when including higher orders. The convergence also depends on an appropriately chosen scale, which is usually set in the range of the momentum transfer during the collision.

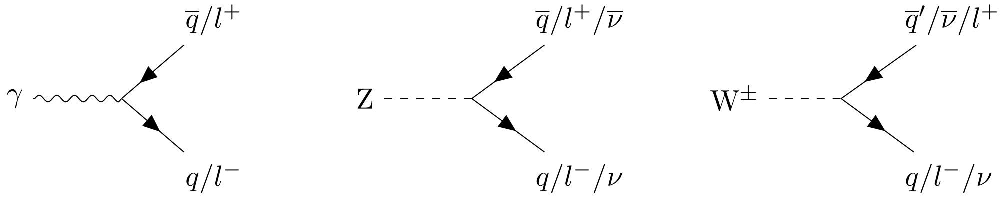
All three fundamental forces can be described within the same mathematical framework, differing only in the consequences following from the gauge symmetry, such as coupling constants and the behaviour of the gauge bosons. These consequences are subject of the next sections.

2.3 Electroweak theory

Quantum electrodynamics (QED) is an abelian gauge theory. The generators of the gauge group U(1) commute, leaving no self-couplings for the single massless boson, the photon. As a consequence, the electromagnetic force has an infinite range with a decreasing field strength, as known from the Coulomb potential.

In contrast to QED, the weak interaction is described by a non-abelian gauge theory. The gauge group SU(2) induces three bosons which were experimentally found to carry masses, the W^+, W^- and the Z boson. The extension of the Lagrangian by simple mass terms contradicts its gauge symmetry, such that masses have to be introduced indirectly via the *Higgs mechanism*. An additional scalar boson that couples to the gauge bosons creates terms that take the role of mass terms, but preserve the gauge symmetry.

The electroweak theory is a unification of weak interaction and QED. The combination of the U(1) and the SU(2) gauge groups introduces two coupling constants and four gauge bosons that all couple to the Higgs boson. Only linear combinations of the gauge bosons are physically relevant and represent the massive W and Z bosons and the massless photon. The linear combination acts like a rotation around the angle Θ_w, called the



(a) photon vertex **(b)** Z boson vertex **(c)** W boson vertex

Figure 2.3: Fermion vertices appearing in electroweak theory. **(a)** A photon γ interacts with charged fermions, i.e. quarks q, antiquarks $\overline{q}$, charged leptons l^- and charged antileptons l^+. **(b)** Apart from quarks and charged leptons, a Z boson also interacts with neutral fermions, i.e. neutrinos ν and antineutrinos $\overline{\nu}$. **(c)** Like a Z boson, a W boson interacts with all fermions, additionally allowing for changing the type (*flavour*) of the fermion. Furthermore, there are boson self-interaction vertices that are not shown here.

Weinberg angle. The angle is given by the relation between the physical masses m^{W} and m^{Z} of the bosons.

$$\cos\left(\Theta_w\right) = \frac{m^{\mathrm{W}}}{m^{\mathrm{Z}}} \tag{2.4}$$

The experimentally derived values of the masses can be found in Figure 2.1.

As a consequence of the relatively large masses of the interaction bosons, the range of the weak interaction is limited, whereas the electromagnetic interaction has infinite range, mediated by the massless photon.

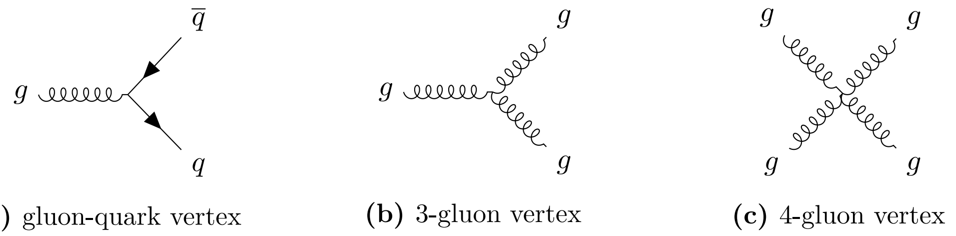
Figure 2.3 gives a summary of the interaction vertices corresponding to the electroweak force that involve fermions. In addition, there are boson self-interaction vertices involving photons, Z and W bosons, resulting from the non-abelian nature of the SU(2) gauge group.

2.4 Quantum chromodynamics

Quantum chromodynamics (QCD) is the quantum field theory based on the SU(3) symmetry describing the strong interaction between particles carrying colour charge. The symmetry introduces one coupling constant g_{S} and eight interaction bosons, called *gluons*. Though gluons are massless, the strong force is known to be delimited to short distances. This is due to additional gluon self-interaction terms in the Lagrangian, resulting from the non-abelian nature of the theory. Figure 2.4 shows the vertices appearing in QCD. In contrast to QED, the coupling constant of QCD decreases with increasing values of the renormalization scale μ_{R}.

At short distances, corresponding to large energies and momenta, strongly interacting particles, i.e. quarks and gluons, behave like free particles, a phenomenon known as *asymptotic freedom*.

On the other hand, the strong interaction increases with small energies, corresponding to large distances. To separate two quarks, energy has to be expended that is stored in the interjacent gluon field. At a certain point, it is energetically favourable for a gluon to



(a) gluon-quark vertex **(b)** 3-gluon vertex **(c)** 4-gluon vertex

Figure 2.4: Vertices appearing in QCD. **(a)** A gluon g interacts with quarks q and antiquarks $\overline{q}$. Gluons interact with each other via **(b)** 3-gluon vertices and **(c)** 4-gluon vertices.

split into a quark-antiquark pair. The effect is that all particles carrying colour charge form colourless bound states, a phenomenon known as *confinement*. Quarks do not occur separately in nature, but always in the form of *hadrons*, either as three-quark (*baryons*, such as the proton) or quark-antiquark combinations (*mesons*). One of the consequences of confinement is the *hadronization* of strongly interacting particles occurring in the final state of a scattering process. This leads to collimated bunches of hadrons, known as *jets*, which are discussed in Section 2.6.

Parton distribution functions

The description of hadrons as a three-quark or quark-antiquark bound state is not complete. Gluons can be emitted by quarks and subsequently split into two or three gluons due to the self-coupling, but also create quark-antiquark pairs of any flavour. Thus, any quark, antiquark or gluon, commonly referred to as *parton*, has a certain probability to be found inside a hadron.

If a proton, as an example for a hadron, is travelling with a momentum p, each parton can be understood to carry the momentum xp, with $0 < x < 1$. The corresponding *parton distribution function* (PDF) $f_a(x)$ contains the probability of finding the parton of type a carrying the momentum fraction x_a. These functions appear when calculating the cross section of processes where hadrons are involved, such as proton-proton scattering. The subprocess cross section $\sigma_{ab}(x_a, x_b)$ of parton a scattering with parton b, where each parton carries the momentum fraction x_a and x_b, respectively, can be calculated perturbatively using Feynman diagrams. The cross section σ_{pp} for proton-proton scattering then follows as a sum of folding integrals of the partonic cross sections with the PDFs.

$$\sigma_{\mathrm{pp}} = \sum_{a,b} \int_0^1 \mathrm{d}x_a \int_0^1 \mathrm{d}x_b \; \sigma_{ab}(x_a, x_b, \mu_{\mathrm{F}}, \mu_{\mathrm{R}}) \; f_a(x_a, \mu_{\mathrm{F}}) f_b(x_b, \mu_{\mathrm{F}}) \tag{2.5}$$

This multiplicative separation of the PDFs from the subprocess cross sections is theoretically justified by the factorization theorem that requires the introduction of a factorization scale μ_{F} on which the PDFs depend. Roughly speaking, μ_{F} defines the separation of the range for the validity of the perturbative treatment of the subprocess calculation and the non-perturbative treatment of the PDFs. Together with the renormalization scale μ_{R}, it is considered in the scale uncertainty which decreases when including calculations at higher orders.

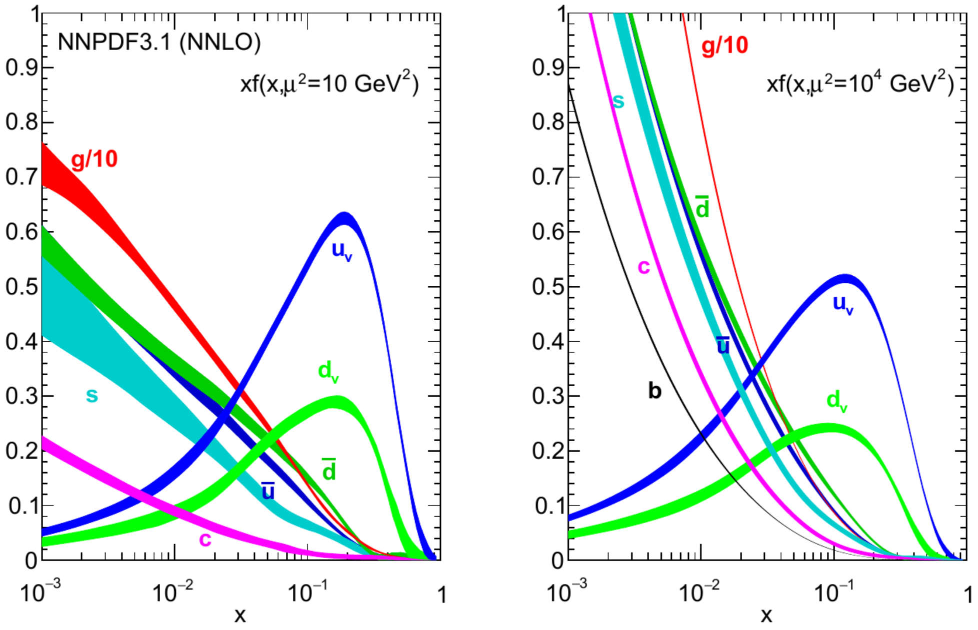


Figure 2.5: NNPDF3.1 parton distribution functions for up (u_v) and down (d_v) valence quarks, strange (s), charm (c) and bottom (b) sea quarks, up ($\bar{\mathrm{u}}$) and down ($\bar{\mathrm{d}}$) antiquarks and the gluon (g) at a scale of $\mu = \sqrt{10}\,\mathrm{GeV}$ (left) and $\mu = 100\,\mathrm{GeV}$ (right), determined by NNPDF [3]. The evolution between the scales is described by the DGLAP equations [12–14].

There is no technique known yet that allows the calculation of PDFs from first principles, such that their determination is left to experiments. For this purpose, measurements are compared to theoretical predictions and the PDFs are fitted such that the best agreement between measurement and prediction is achieved. The fit result depends on the scale μ of the measurement, which is usually taken as both the factorization scale μ_F and the renormalization scale μ_R, and is itself commonly chosen in the range of the momentum transfer during the collision. The evolution of the PDFs to other scales can be calculated by the *Dokshitzer-Gribov-Lipatov-Altarelli-Parisi* (DGLAP) equations [12–14].

There are several groups that provide proton PDFs, each differing in the selection of experiments, the PDF parametrization or the method of the fitting procedure. Figure 2.5 shows the NNPDF3.1 fit results for PDFs estimated by the NNPDF collaboration at two different scales $\mu = \sqrt{10}\,\mathrm{GeV}$ and $\mu = 100\,\mathrm{GeV}$ [3].

Interpolation techniques

To evaluate the integrals in Equation 2.5, the subprocess cross sections σ_{ab} need to be known at any values of (x_a, x_b). In practice, this approach requires an enormous amount of computing power and is therefore not feasible. Alternatively, techniques have been developed which evaluate the subprocess cross sections and the PDFs at a number of grid values of x and interpolate the intermediate behaviour. The FASTNLO project is an example for an implementation of this technique [15].

The interpolated results for the subprocesses are stored in interpolation grids that enable a flexible use of different PDFs as well as the variation of the renormalization and factorization scales or the value of the strong coupling constant. The difference between cross section predictions obtained with and without interpolation is aimed to be less than 0.1 %, requiring the interpolation nodes to be chosen reasonably. The method is particularly important when performing PDF fits that require many iterations of changing PDFs and comparing the results to experimental measurement.

2.5 Monte Carlo simulations

With the theoretical framework at hand, predictions for cross sections of a scattering process of interest can be calculated. Those degrees of freedom that are not of particular interest (such as spin or kinematic properties) or even impossible to probe (such as colour charge) are averaged over all possible realizations in the initial state and summed or integrated over the possibilities in the final state. As the integration phase space can have large dimensionality, a Monte Carlo integration method is the tool of choice.

Particles in the initial state are usually understood to have well-defined momenta. For hadrons in the initial state, this statement holds only for the hadron as a whole, though its partons follow the respective PDFs. Particles in the final state can have any kinematic properties which energy conservation renders possible.

The cross section of a process can be derived perturbatively at a fixed order by the use of calculation frameworks such as MADGRAPH5 and aMC@NLO [17] or NNLOJET [18].

However, an accurate comparison to an experiment requires a simulation under consideration of additional effects that are explained in the following. Such a simulation is performed using Monte Carlo techniques, implemented by event generators such as PYTHIA8 [19], HERWIG++ [20, 21] or Sherpa [22].

The dominant effects in the context of proton-proton scattering are *multiple parton interaction*, *parton showering* and *hadronization*.

- **Multiple parton interaction**: In a proton-proton collision there is usually one so-called *hard* scattering process, meaning a parton-parton collision with high momentum transfer. Apart from this hard process, the remnants of the protons experience *multiple parton interaction* (MPI) and form the *underlying event* (UE) [23].
- **Parton shower**: Both initial state and final state partons can spontaneously emit other partons that contribute to the event in form of additional jets. The main challenge of parton showering is the correct use in conjunction with real emission contributions to the hard process. To avoid double-counting, algorithms have been developed to combine parton showers with LO or NLO matrix elements (CKKW [24], MLM [25], FXFX [26], POWHEG [27], MC@NLO [28]).
- **Hadronization**: Due to confinement, partons in the final state can not exist separately, but fragment into a multitude of partons that form colourless hadrons.

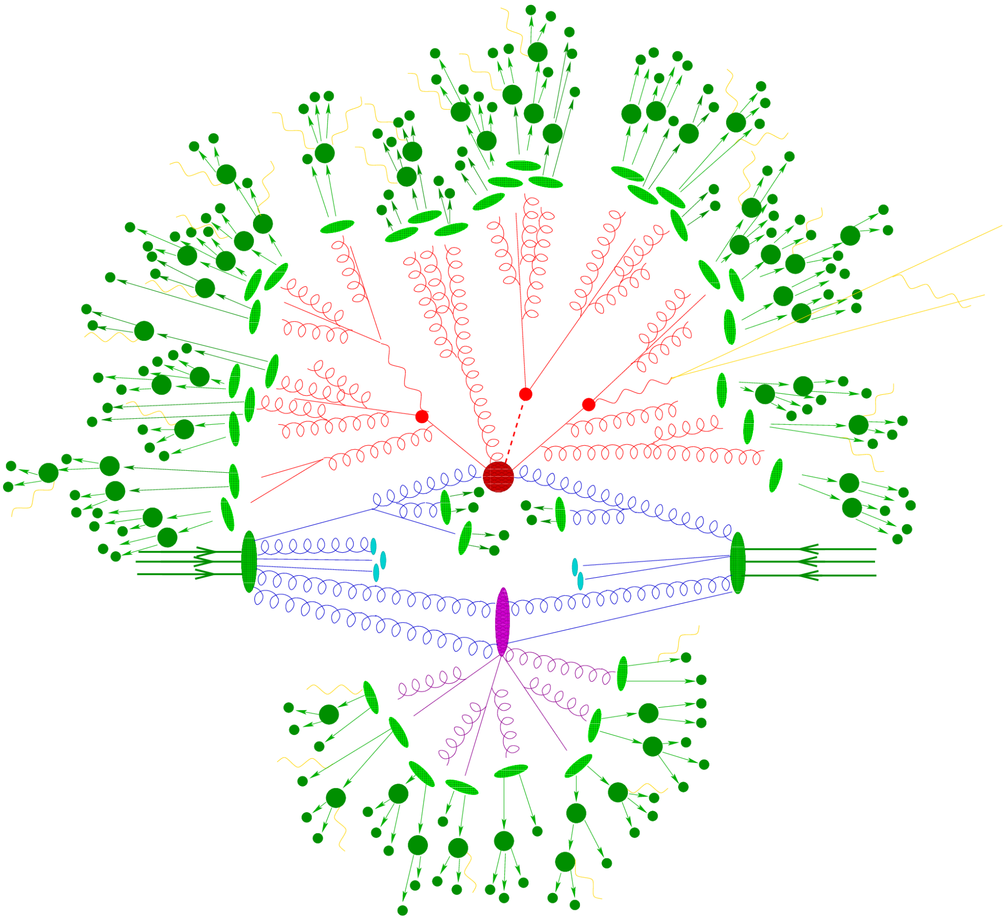


Figure 2.6: Sketch of a simulated proton-proton collision including all effects covered by event generators. The dark red blob in the centre illustrates the hard scattering between two partons, resulting in a number of outgoing particles whose further decay is illustrated by the bright red blobs. Parton showers can emerge from partons before (initial state radiation (ISR), blue lines) as well as after the collision (final state radiation (FSR), red lines). The underlying event consists of remnants of the proton (turquoise blobs) that can also experience multiple parton interactions illustrated by the purple blob and lines. The bright green blobs indicate the point where hadronization of the outgoing partons sets in, resulting in the multitude of particles illustrated by dark green blobs. Furthermore, any charged particle can emit soft photons (i.a. bremsstrahlung). Taken from [16].

> Hadronization takes place at scales where the perturbative approximation is not applicable. The used hadronization models differ between the event generators. Considering the interactions between particles carrying colour charge to be proportional to their distance leads to the *Lund string model* [29], which is chosen for Pythia8. A hadronization model based on the formation of color singlets organized in clusters, a phenomenon called *preconfinement* [30], is preferred in Herwig++.

Furthermore, electroweak interactions can cause soft photon radiation emitted by charged particles at any stage of the simulation. Figure 2.6 illustrates the interplay of the previously described effects, eventually resulting in an event with a multitude of final state particles.

Each of those effects are implemented using parameters that are tuned such that the result best describes the experimental data. There is no certainty which choice of simulation method is better, each with its own advantages and disadvantages in its predictive power. The agreement to experimental results has to be validated for every measurement, disagreements have to be considered by systematic uncertainties.

As a last step, the interaction of the particles in the final state with the detector can be simulated with Geant4 [31, 32].

2.6 Jets

As a result of confinement, partons in the final state of a process fragment into a multitude of particles, approximately moving into the same direction. Algorithms have been developed to cluster the four-momenta of particles unambiguously into a well-defined *jet*, whose four-momentum serves as a measure for the original parton before the fragmentation. Jets are among the most common observable objects in collider physics.

As jet clustering combines four-momenta of particles, the coordinate system widely used in collider physics is discussed in the following.

2.6.1 Coordinate conventions

In a typical scattering process, a beam of particles is incident either on a fixed target or head-on a second beam of particles. Both scenarios inherit a cylindical symmetry with the beam axis as the centre line. It is therefore commonly chosen as the z-axis, the transverse component parallel to the ground is commonly chosen as the x-axis.

A particle leaving the collision point with a momentum $\vec{p} = (p_x, p_y, p_z)$ can then be described by its transverse momentum $p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}$, the angle ϕ between the x-axis and the projection on the x-y-plane and the angle θ between $\vec{p}$ and the z-axis.

The relativistic nature of the particles involved in high energy physics motivates the transformation of the angle θ to the *pseudorapidity* η, which can also be written using $\vec{p}$ and p_z.

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) = \frac{1}{2}\ln\left(\frac{|\vec{p}| + p_z}{|\vec{p}| - p_z}\right) \tag{2.6}$$

The latter notation is the reason for its denotation, as it is closely related to the *rapidity* y along the beamline.

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right) \tag{2.7}$$

For particles whose masses can be neglected compared to their momentum, which is often the case for decay products in particle physics, the rapidity and the pseudorapidity are identical.

Differences in the rapidity y and in the angle ϕ are invariant when transforming the laboratory system to a reference system *boosted* along the beamline. By the use of these coordinate conventions, the distance between particles with the four-momenta $p_i = (E_i, \vec{p}_i)$ and $p_j = (E_j, \vec{p}_j)$ in the y-ϕ-plane can be expressed by Δ_{ij}.

$$\Delta_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2} \tag{2.8}$$

Δ is invariant under boosts along the beamline and can therefore be used by algorithms that identify sets of particles most likely to originate from the same parton.

Analogously, the *angular* distance ΔR between particles is defined by replacing the rapidity y by the pseudorapidity η in the definition of Δ in Equation 2.8.

2.6.2 Clustering algorithms

There are several requirements imposed on an algorithm to guarantee its applicability to general particle physics. First, it needs to be applicable to experimental measurements as well as to theory predictions. Regarding the theory predictions, the algorithm must be capable of dealing with any simulation level such as *parton level*, *particle level* or *detector level*.

- At **parton level**, the event topology includes partons and non-strongly interacting particles in the final state. Only effects that can be described in the perturbative approximation, such as parton shower, contribute to the parton level. Parton-level jets typically consist of a small number of constituents.

- The **particle level** represents the outcome of a scattering process after consideration of non-perturbative effects such as hadronization. Partons are fragmented into cascades of hadronic particles that are colour-neutral. The comparison of the particle level to the parton level can be used to derive non-perturbative (NP) corrections on cross sections calculated at fixed order. Particle-level jets usually consist of numerous constituents.

- To reach the **detector level**, the particle-level event undergoes a detector simulation, which makes it finally possible to compare the simulation to an actual measurement. Since the detector level is the only level actually accessible by experiments, the simulations can be used in order to correct measurements for detector effects. The constituents of detector-level jets are four-momenta which are reconstructed from energy depositions in the detector.

Jet algorithms can be classified either as *cone algorithms* or as *sequential recombination algorithms*. Cone algorithms sum up all four-momenta within a certain distance Δ, as defined in Equation 2.8, from a starting four-momentum to calculate the jet four-momentum. Sequential recombination algorithms use an iterative approach. A distance measure is defined that can take into account the transverse momentum in addition to angular and rapidity information of particles.

A widely used distance measure d_{ij} between the four-momenta p_i and p_j includes a parameter p, a distance parameter R, also referred to as the jet size, and the distance Δ_{ij} as defined in Equation 2.8.

$$d_{ij} = \min\left(p_{\mathrm{T}i}^{2p}, p_{\mathrm{T}j}^{2p}\right) \frac{\Delta_{ij}^2}{R^2} \tag{2.9}$$

Furthermore, the distance of a four-momentum p_i to the beam axis is defined as d_{iB}.

$$d_{iB} = p_{\mathrm{T}i}^{2p} \tag{2.10}$$

A sequential recombination algorithm based on this distance measure creates a list of all distances d_{ij} by combining all pairs of four-momenta in the event, extended by the list of distances d_{iB} of all four-momenta to the beam axis. It merges particles by summing up the pair of four-momenta with the smallest distance d_{ij} and recalculating all distances. If the smallest distance in the list is d_{iB}, a termination condition is reached and the four-momentum p_i is called a jet. All entries involving the four-momentum p_i are removed from the list of distances. The algorithm continues until no entry is left.

The choice of the parameter p defines further properties of the algorithm. A widely used example of a sequential recombination algorithm is the anti-k_t algorithm based on this distance measure with $p = -1$ [33].

A jet area can be assigned to each jet by artificially adding infinitesimal four-momenta to the event, which are uniformly distributed in the y-ϕ-plane. The jet area is defined as the domain covered with those infinitesimal four-momenta which are clustered into the jet. In case of the anti-k_t algorithm, it can be understood as the area of the cone base that contains the eventual jet, having a cone base radius R.

This definition of a jet area is possible only if the algorithm fulfils *infrared and collinear safety*, which is one of the most important requirements on jet algorithms. Because partons are non-observable objects, cross sections are typically calculated with jets instead of partons in the final state. Collinear or infrared emissions of partons correspond to real contributions in higher order calculations. If the jet clustering is not sensitive to these emissions the algorithm is called infrared and collinear *safe* and can be used within perturbation theory [34, 35].

The application of such an algorithm always results in the same, well-defined set of clustered jets. However, since the mapping of a particular four-momentum onto a jet depends on the chosen clustering algorithm, the topology of jets in an event is ambiguous. The *out-of-cone* effect occurs if a particle that actually originates from the parton is not covered by the jet area. Conversely, particles originating from other partons can be included by the clustering. Thus, the choice of the jet algorithm is crucial for the calculation as well as for the measurement of cross sections of proton-proton scattering processes.

CHAPTER 3

Experimental setup

The European Organization for Nuclear Research, CERN, situated at the border between France and Switzerland close to Geneva, is nowadays one of the most prominent research institutions in the world. Throughout the last 60 years, it has had a huge impact on the technological progress of physical research, eventually leading to the most powerful particle accelerator ever built, the Large Hadron Collider (LHC).

This progress was driven by the intention to test theoretical predictions derived from the Standard Model of particle physics. Scattering experiments procure a controlled environment, in which the interactions between particles can be described by the Standard Model. To probe particles carrying high masses or interacting at small distances, high energies are necessary which, in a controlled environment, can only be provided by the LHC.

3.1 The Large Hadron Collider

The LHC consists of a circular tunnel with a circumference of 27 km in a mean depth of 100 m below ground, populated with superconducting electromagnets whose coils contain an electric current which is able to create a magnetic field of up to 8 T. These fields keep protons (or heavy ions) on track at a velocity close to the speed of light. The protons originate from ionized hydrogen and undergo several stages of acceleration before they eventually enter the LHC, arranged as bunches, each containing billions of protons. Since the year 2015, they accumulate an energy of $\sqrt{s} = 6.5\,\mathrm{TeV}$ during the acceleration. Figure 3.1 illustrates the LHC complex, in which several experiments are embedded as well. Two beams consisting of thousands of bunches are travelling in opposite directions and are forced to cross at four interaction points where the main experiments are situated. The number of protons contained in each beam divided by the circulation time of about $100\,\mu$s and the beam area gives rise to the luminosity, which is a measure for the particle flux. In principle, a collision is possible every 25 ns, when two bunches cross each other. If a collision occurs, detectors will observe the remnants of the scattering process. One of the detectors at the LHC is the Compact Muon Solenoid (CMS), which recorded the data used within this thesis.

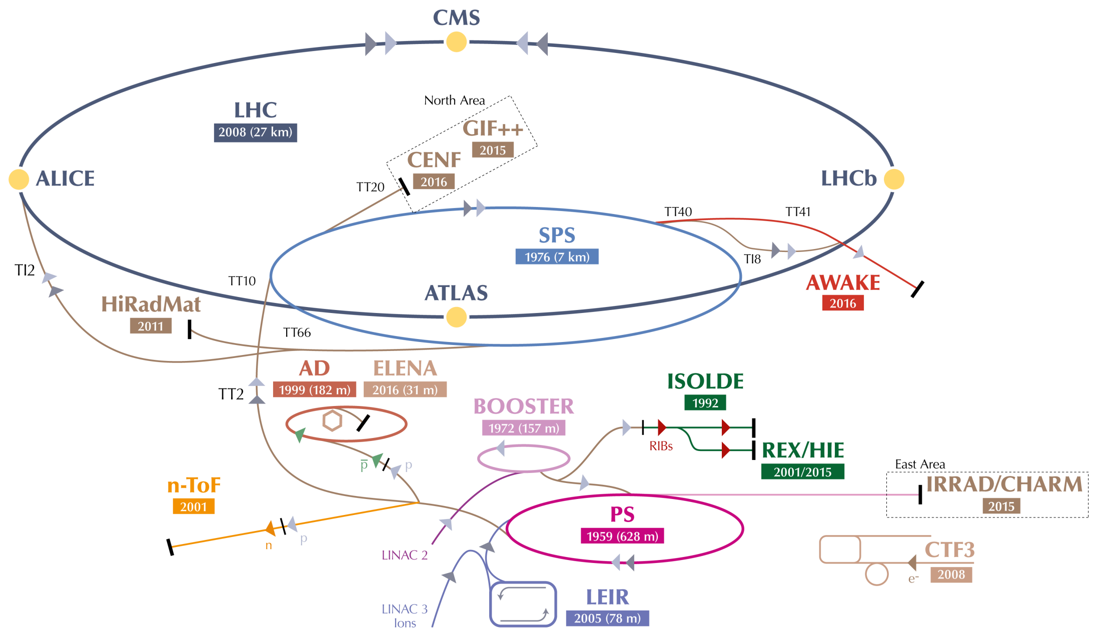


Figure 3.1: The LHC complex at CERN close to Geneva. Besides the main experiments CMS, LHCb, ATLAS and ALICE, numerous other experiments are embedded. The way of the protons starts at the linear accelerator LINAC2, passes the pre-accelerators BOOSTER, PS (Proton Synchrotron) and SPS (Super Proton Synchrotron) and finally leads to the LHC. Illustration taken from [36].

3.2 The Compact Muon Solenoid

The CMS experiment is placed in a cavern at one of the four interaction points within the LHC tunnel. It is designed as a general-purpose detector to study a broad range of physical processes. The detector composition of several layers of detector subsystems is illustrated in Figure 3.2. Its name reflects the compact way to arrange most subsystems inside a solenoid magnet and the high performance for detecting muons.

3.2.1 Detector subsystems

Each detector subsystem plays a specialized role in the identification of scattering products. In the following, these subsystems are shortly described, starting with the innermost parts. A detailed description of the detector can be found in [38].

- **Silicon tracker:** The tracking system is composed of several layers of silicon pixels (inner tracker) and silicon strips (outer tracker). Charged particles passing a pixel or a strip create free charge carriers in the semiconductor material. As a result, an applied voltage creates an electric current which can be read out and detected

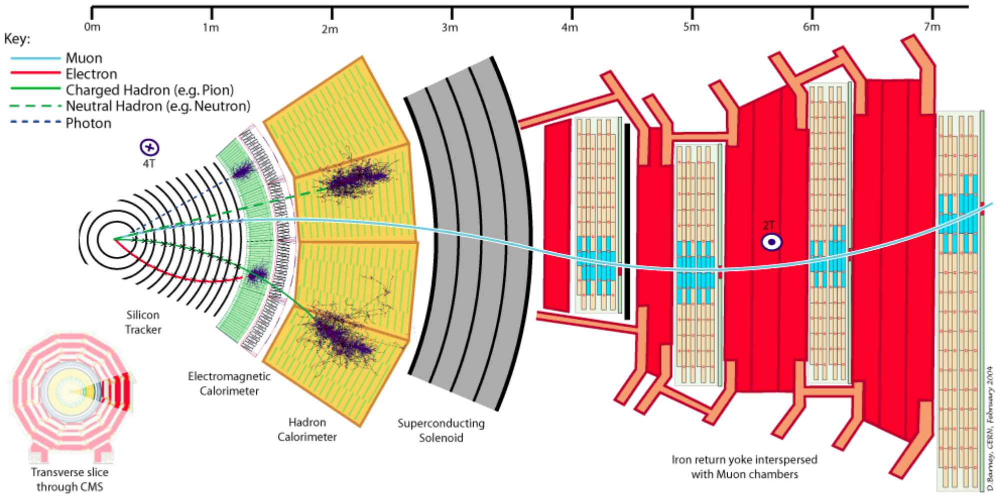


Figure 3.2: The CMS detector. A small piece of a transverse slice through the CMS detector exhibits the subsystem structure. The silicon tracker detects trajectories of charged particles, the calorimeters measure particle energies. The trajectories are bent because of the superconducting solenoid magnet, whose field is extended to the outer region by an iron return yoke. The muon chambers detect muon trajectories. The combination of tracks and energy depositions in the calorimeters allows the classification of particles as electrons, photons, muons, charged and neutral hadrons. Illustration taken from [37].

as a hit. These hits are combined to reconstruct particle tracks. Their curvature provides information about the charge and the momentum of the particle.

- **Electromagnetic calorimeter (ECAL):** Electromagnetically interacting particles such as photons, electrons and positrons are absorbed in lead tungstate scintillators. The emerging electromagnetic showers are measured by photodetectors and provide information about the particle energies.

- **Hadron calorimeter (HCAL):** Due to their higher radiation length, hadrons pass the ECAL and enter an alternating arrangement of brass and steel absorbers and plastic scintillators. The hadrons create cascades of further particles in the absorbers, whose energies are measured via the scintillators. The thickness of this arrangement is of the order of multiple absorption lengths, ensuring that nearly all particles created in the collision can be detected.

- **Superconducting solenoid:** A superconducting solenoidal coil contains a current whose magnetic field of 3.8 T is responsible for the curved paths of charged particles. The field is stabilized by an iron return yoke outside of the electromagnet.

- **Muon chambers:** Muons are particles which are able to escape all layers of the detector. They leave tracks in the muon chambers as well as in the silicon tracker. There are three types of muon detectors that all rely on the ionization of a gas volume by a passing muon. *Drift tubes* (DT) contain a positively charged stretched wire that detects the drift of ionized gas particles. *Cathode strip chambers* (CSC) use the same mechanism, supplemented by cathode strips perpendicular to the anode wires, providing additional position information. *Resistive plate chambers* (RPC) consist of two high resistivity plastic plates separated by the gas volume and are particularly used for triggering.

The subsystems are arranged around the collision point, separated into a central and a forward part, such that particles can be detected in a wide pseudorapidity range, excluding those travelling with a small angle along the beam line. All the information provided by the subsystems is read out by electronics and handed over to a two-staged trigger system that decides for each collision independently whether or not it is worth to be stored.

3.2.2 Trigger System

Two proton bunches cross each other at a rate of 40 MHz in the centre of the detector, giving rise to one potential collision each 25 ns. Since most of these collisions only happen at low energy scales, their physical relevance is negligible. The disk space required for one reconstructed event is in the range of 1 MB, meaning that about 40 PB of data would be stored per second if every event was kept, more than even the combined computing power of the CERN infrastructure could handle. Therefore, a two-staged trigger system scans each event for potentially interesting features, reducing the rate of data flow to about 1 GB per second [39].

- **L1:** The level 1 trigger is mainly hardware-based to allow for triggering at the high rates necessary at the LHC. Only simple signatures such as an energy deposition in a calorimeter or a track in the muon system are taken into account for a fast selection.
- **HLT:** The software-based high level trigger preliminarily reconstructs events that pass the L1 selection. A trigger computing farm above ground is necessary to evaluate the information from the different subsystems in short timescales.

There is a large number of different trigger requirements, also called *trigger paths*, combining L1 and HLT requirements used for a classification of the event, ranging from the detection of a single particle above an energy threshold to the sum of all detector components. If any of these paths is triggered, the event will be classified accordingly.

The remaining data are stored at computing clusters at CERN and around the world, available for further analysis.

3.2.3 Computing infrastructure

The worldwide LHC computing grid (WLCG) manages the distribution of the data collected by all LHC installations to computing resources around the world [40]. There are several hierarchical levels within its organisation. The CERN Data Centre at the

CERN site, also denoted by the name *Tier 0*, provides internal storage for raw data collected by the experiments and computing power for first offline processing steps.

Currently, 13 Tier 1 centres are distributed at universities and institutes around the world, of which one is situated in Karlsruhe. They provide computing power mainly for centrally coordinated data reprocessing and simulation. Additionally, they store copies of the experimental data accessible for researchers.

More than 100 Tier 2 centres, also distributed at universities and institutes around the world, provide computing power for data analysis by individual working groups, as well as for simulations that are eventually stored and made available at the Tier 1 centres.

Furthermore, several computing centres that are available for the local research groups are categorized as Tier 3. At KIT, additional computing resources that are not exclusively meant for particle physics purposes are integrated on-demand. One of them is the high performance bwForCluster NEMO at the university of Freiburg, which was used for most of the analysis in the scope of this thesis [41].

3.3 Reconstruction of physical objects

3.3.1 Particle flow algorithm

Besides the composition of the CMS detector, Figure 3.2 also illustrates how information from all subsystems is combined to reconstruct physical objects. The method used within the CMS collaboration is called the *particle flow* (PF) algorithm and leads to a collection of PF candidates which are categorized as one of five different types [42, 43].

- **Muon candidate:** A trajectory within the muon system only, combining information from DT, CSC and RPC, defines a *standalone* muon, whereas a track reconstructed within the silicon tracker, whose extrapolation matches at least one hit in the muon system, defines a *tracker* muon. If a standalone muon trajectory matches a track, it is stored as a *global* muon. Furthermore, the sign of the track curvature provides information about it being a muon or an antimuon [44].

- **Electron candidate:** The remaining tracks within the silicon tracker are extrapolated to the calorimeters. A track which is compatible with an energy deposition in the ECAL defines an electron or a positron, depending on its curvature.

- **Photon candidate:** An energy deposition in the ECAL above a given threshold without any compatible track defines a photon.

- **Charged and neutral hadron candidates:** In the same manner, the remaining tracks are compared with depositions in the HCAL to define charged and neutral hadrons.

If a PF candidate is reconstructed, all depositions and hits used for its reconstruction are removed from the event. Mathematically, one event is built out of the set of the four-momenta of all PF candidates.

Missing transverse momentum

The resulting PF candidates will further be used for jet clustering and tau lepton reconstruction, but also for the computation of yet another important quantity: the *missing tranverse momentum* $\vec{p}_\mathrm{T}^\mathrm{miss}$. It is defined as the negative transverse component of the momentum sum of all reconstructed PF candidates. Its importance results from the initial state of the colliding particles, which does not have any momentum component perpendicular to the beam direction. Thus, $p_\mathrm{T}^\mathrm{miss}$ would vanish if *all* particles emerging from the collision were measured correctly. It can therefore be used for two important tasks: calibration of the detector using events whose properties are well understood, such as Z boson events, where the Z boson decays to a muon-antimuon or an electron-positron pair, or for an indirect detection of particles that are invisible to the CMS detector, such as neutrinos or any exotic kind of unknown (dark) matter.

3.3.2 Pileup

As protons travel through the LHC in the form of bunches, it is inevitable that the products of the main proton-proton collision are overlaid by remnants from additional proton-proton collisions. *Pileup* is a collective name for any depositions in the detector that bias the event reconstruction by not originating from the main collision. As a result, additional jets can be reconstructed, which typically carry low energies. There are several observables which quantify the pileup contribution of an event, all of which increase with the LHC operating at higher luminosities.

- Tracks of charged particles enable the reconstruction of *primary* vertices as their crossing points along the beamline [45]. The number n_PV of primary vertices serves as a measure for pileup.

- A second quantity that can serve as a measure for pileup is the p_T-density ρ. Assuming that pileup is isotropically distributed in an event, it can be estimated as the median value of all jet transverse momenta divided by the respective jet area [46]. However, ρ is also biased by contributions from the underlying event and detector noise.

- In simulations, pileup contributions are manually added to each event, such that the actual number of pileup interactions n_PU is well known. In the experiment, this number can only be estimated averaged over a certain time interval, a *luminosity section*, during which the beam properties can be regarded as constant. The luminosity integrated over a luminosity section, which is approximately 24 s, multiplied by the total proton-proton scattering cross section gives rise to the mean number of pileup interactions $\mu = \langle n_\mathrm{PU} \rangle$.

Pileup mixing in simulations suffers from the fact that the actual pileup contribution in the experiment is not known in advance. To gain more flexibility, each simulated event can be weighted, so that the distribution of the pileup quantity in the simulation matches the one in the measurement. The official recommendation is to use μ for pileup reweighting.

3.3.3 Muon reconstruction

In addition to the PF algorithm, further criteria can be applied to improve the muon reconstruction. To reject particles falsely categorized as muons, identification working points are defined, the most common being the *loose* and *tight* working points [44].

- **loose:** Any muon reconstructed as a global or a tracker muon by the PF algorithm passes the loose working point, maximizing the muon reconstruction efficiency.
- **tight:** For a muon to pass the tight selection, in addition to the reconstruction by the PF algorithm as a global and a tracker muon, criteria as summarized in Table 3.1 are applied, minimizing the possibility to misidentify a track for a muon.

Table 3.1: Criteria for a muon to pass the *tight* identification as applied in 2016.

$\chi^2/n.d.f.$ of the global muon track fit	< 10
Number of hits in muon system included in the global muon track fit	≥ 1
Number of stations (i.e. layers) hit in muon system	≥ 2
Number of layers hit in the pixel or strip tracker	≥ 6
Number of layers hit in the pixel tracker	≥ 1
Transverse distance $\lvert d_{xy}\rvert$ of the track to primary vertex	$< 0.2\,\mathrm{cm}$
Longitudinal distance $\lvert d_z\rvert$ of the track to primary vertex	$< 0.5\,\mathrm{cm}$

Furthermore, to discriminate between muons originating from heavy-flavour decays and muons from boson decays, the muon isolation is used, taking into account particles within a surrounding cone. The PF relative isolation is defined as the ratio of the sum of the transverse momenta of photons, charged and neutral hadrons with an angular distance of less than $\Delta R = 0.4$ to the transverse momentum of the muon. The *loose* and the *tight* working points reject muons with a relative isolation of more than 25 % or 15 %, respectively.

3.3.4 Jet reconstruction

Within the CMS collaboration, jets are clustered from PF candidates using a jet clustering algorithm of choice. The anti-k_t algorithm is most commonly used with a distance parameter of $R = 0.4$ or $R = 0.8$ [33]. The distance parameter $R = 0.4$ is applied for jets used within this thesis.

To reject noise or badly reconstructed jets, identification working points are defined which take into account energy fractions of PF candidate types and their multiplicities, depending on the pseudorapidity η of the jet. The selection criteria of the *loose jet ID* working point are given in Table 3.2 [47].

Table 3.2: Criteria for a jet with $|\eta^{\mathrm{jet}}| < 2.4$ to pass the *loose* identification as applied in 2016.

Neutral Hadron Fraction	< 0.99
Neutral EM Fraction	< 0.99
Number of Constituents	> 1
Charged Hadron Fraction	> 0
Charged Multiplicity	> 0
Charged EM Fraction	< 0.99

Pileup creates a significant number of detector depositions that are part of the collection of PF candidates contributing to the jet reconstruction. To identify jets which are clustered from pileup contributions, a supplemental identification criterion is provided as *loose*, *medium* and *tight* working points. This *pileup jet ID* is based on a multivariate analysis of a multitude of parameters describing the shape and the vertex origin of the constituent particles of a jet [48].

There are additional methods to reduce the pileup contribution which is as well inherent to every jet that does not completely originate from pileup. The *charged hadron subtraction* (CHS) method vetoes PF candidates that are identified not to originate from the main collision vertex. As this method does not affect the contribution of neutral particles, the *pileup per particle identification* (PUPPI) method independently attaches a weight to each PF candidate before jet clustering. The weight reflects the probability for the PF candidate to originate from pileup and is derived based on a multivariate analysis [49].

At last, corrections for the jet transverse momentum are used to assign the reconstructed transverse momentum on detector level to the one on particle level. Chapter 4 gives an overview about how these corrections are derived by jet energy calibration.

CHAPTER 4

Jet energy calibration using Z ($\rightarrow \mu\mu$) + jet events

Jets are among the most common observable objects in collider physics. Many analyses rely on a precise estimation of the energy of a jet to determine cross sections involving outgoing partons.

Confinement forces a parton leaving the collision point to transform to a collimated bunch of hadronic particles that form a jet, a process which is referred to as hadronization. Jet clustering combines the four-momenta of these particles to derive a measure for the original parton.

On detector level, inefficiencies and noise in the detector components, as well as additional contributions from pileup bias the jet momenta. In addition, any mismodelling in the detector simulation can have similar effects. For these reasons, the jets need to be calibrated to achieve accurate results.

In the following, the jet energy calibration is discussed, as carried out by the CMS collaboration. The calibration workflow is separated into multiple stages which provide jet energy corrections (JEC) on different aspects of jet mismeasurement. The used methods have been developed for the data taking period between of 2010 to 2012 with the LHC operating at a centre-of-mass energy of 7 TeV [50] and 8 TeV [51] (Run I). A short description of the stages and methods is provided in Section 4.1.

The analysis of Z boson events in association with jets, where the Z boson decays to a muon-antimuon pair (Z ($\rightarrow \mu\mu$) + jets), is discussed in Section 4.2 as an example for the estimation of *residual* corrections, which account for differences between simulation and measurement. This calibration channel and its implementation for the data taking at 13 TeV is part of the work presented in this thesis.

For the final calibration, results of all calibration channels are combined in a global fit, as presented in Section 4.3. Closure tests performed within the scope of this thesis contribute to the validation of the global fit results, as described in Section 4.4. At last, uncertainties on the derived corrections are estimated, which is explained in Section 4.5.

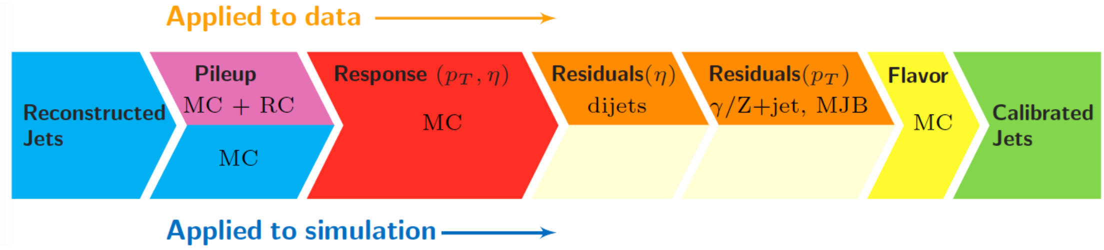


Figure 4.1: Stages of jet energy calibration. Taken from [51].

4.1 Stages of jet energy calibration

JECs are derived in multiple stages that are illustrated in Figure 4.1, using inputs from numerous simulated as well as experimental results. Each stage yields multiplicative factors which are applied consecutively on the jet transverse momentum. The stages are shortly discussed in the following [50, 51].

4.1.1 Pileup offset corrections

The first stage is an offset correction for contributions from pileup and detector noise. Two methods, which make use of pileup observables introduced in Section 3.3.2, have been developed to determine correction factors:

- **Jet area method:** Under the assumption that pileup and noise are isotropically distributed, an offset correction for each jet can be determined by multiplying its area with the p_{T}-density ρ. However, the contributions from the underlying event have to be estimated from no-pileup events and subtracted from the correction because they form an immanent feature of the main proton-proton collision.

- **Average offset method:** Alternatively, correction factors can be derived from the number of primary vertices n_{PV} of the event and an η^{jet}-dependent average offset correction. The latter is derived by scanning for energy depositions in the η-ϕ-plane by randomly placed cones (*random cone* (RC) method) and averaging over ϕ and the events.

In CMS, both methods are combined to a *hybrid method*. The correction dependends on η^{jet}, the jet area and the p_{T}-density ρ of the event.

4.1.2 Simulated response corrections

Correction factors that connect the p_{T} of the detector-level jet to the p_{T} of the particle-level jet are derived from simulation. At first, both jets have to be assigned to each other by *matching*: for each detector-level jet, the particle-level jet with the closest angular distance ΔR, which must not exceed half the jet distance parameter R, is chosen. The ratio of the detector-level to the particle-level transverse momentum is taken as *response* variable $\mathcal{R}$. Averaging over all jets in a certain acceptance range in all events leads to the response as a function of p_T and η of the detector-level jet. The inverse response then serves as a correction factor dependent on the p_T and η of the jet.

4.1.3 Residual corrections

Simulated response corrections are crucially dependent on the simulation input, especially on the detector simulation, but also on the physics modelling used by the Monte Carlo generators. If their modelling differs from nature, additional corrections or uncertainties have to account for these effects.

The determination of the residual corrections is based on the estimation of response variables that compare the p_T of a jet to the p_T of a reference object. The corrections are named *relative* and *absolute* corrections and are derived in two steps.

Relative residual corrections

Relative residual corrections aim at calibrating the p_T of a jet in a high-$|\eta|$ (forward) region in relation to a reference object in the low-$|\eta|$ (central) region. Although other objects are possible, the reference object is chosen to be another jet because of the abundance of dijet events at the LHC. The relative response values derived from the forward jet with respect to the central one are estimated as a function of η of the forward jet.

Absolute residual corrections

Absolute residual corrections aim at calibrating the p_T of a jet in the central region in relation to a well-measured reference object. This object is typically a Z boson, either decaying to a muon-antimuon (Z ($\rightarrow \mu\mu$) + jet) or an electron-positron pair (Z ($\rightarrow ee$) + jet), or a photon (γ + jet). Above a certain $p_\mathrm{T}^\mathrm{jet}$ threshold, a sufficient amount of data is not available using the former processes. A multijet topology, meaning the sum of several already well-calibrated jets, is used as reference object for one high-energetic jet. The response values are determined as a function of p_T of the jet. The methods are combined to cover as large as possible a range in p_T and to reduce the uncertainties of the obtained correction factors.

By combining relative and absolute residual corrections, jets can be calibrated for differences between data and simulation over a wide range of their p_T and η. The work performed within this thesis has made a contribution to the determination of absolute residual corrections by analyzing Z ($\rightarrow \mu\mu$) + jet events. The used methods are explained in more detail in Section 4.2.

4.1.4 Flavour corrections

Jets originating from partons that differ in flavour (flavour in this context refers to gluons as well as to quarks) experience differences in jet clustering. For example, gluons hadronize into a larger number of particles than quarks; bottom or charm quarks decay in association with neutrinos. These differences bias the resulting $p_{\mathrm{T}}^{\mathrm{jet}}$. Therefore, additional corrections are derived from simulations by comparing momenta of jets of different flavour.

Applying flavour-dependent JEC remains an optional task for most analyses. However, the study of jets of different flavour is included in the JEC uncertainties, as described in Section 4.5.

4.2 Absolute residual corrections using Z ($\rightarrow \mu\mu$) + jet events

The analysis of Z + jet events forms an important contribution to jet energy calibration because of the precision that can be achieved. A Z boson decaying to an electron-positron or a muon-antimuon pair can be reconstructed with high accuracy and low background contamination, and is therefore a suitable reference object for the estimation of residual corrections. Especially the precisely measured muon momenta render the Z ($\rightarrow \mu\mu$) + jet topology the most accurate channel for jet calibration.

4.2.1 Methods of absolute residual corrections

The determination of absolute residual corrections is based on the comparison of a jet in the central detector region to a suitable reference object, which in this case is the Z boson. There are two complementary methods that yield a response variable $\mathcal{R}$ to be used for calibration: the p_{T} *balance* method and the *missing transverse momentum (p_{T}^{miss}) projection fraction* (MPF) method.

Transverse momentum balance method

The p_{T} balance method is the straightforward way to obtain a response variable $\mathcal{R}$ by taking the ratio of the transverse momentum of the jet to the one of the Z boson.

$$\mathcal{R}_{\mathrm{p_T bal.}} = \frac{p_{\mathrm{T}}^{\mathrm{jet}}}{p_{\mathrm{T}}^{\mathrm{Z}}} \tag{4.1}$$

In an exclusive Z + jet event topology, with exactly one jet, $p_{\mathrm{T}}^{\mathrm{jet}}$ and $p_{\mathrm{T}}^{\mathrm{Z}}$ must be equal and therefore $\mathcal{R}_{\mathrm{p_T bal.}} = 1$ if the jet was reconstructed correctly. Otherwise, the inverse value of $\mathcal{R}_{\mathrm{p_T bal.}}$ immediately gives rise to a correction factor for the jet, under the assumption that the Z boson is reconstructed correctly. However, the inequality of their transverse momenta is not only induced by detector imperfections, noise and pileup contributions, but also by *additional jet activity.* As additional jets can arise from parton showers or gluon emissions, an exclusive Z + jet event topology is unlikely. To obtain a correction factor from the p_T balance method, the additional jet activity has to be eliminated.

MPF method

An alternative way to obtain a response variable is to take the missing transverse momentum $\vec{p}_\mathrm{T}^\mathrm{\,miss}$ into account. In an exclusive Z + jet event topology and respecting the vector properties of transverse momenta, the $\vec{p}_\mathrm{T}$ of the correctly reconstructed jet and the Z boson must compensate each other.

$$\vec{p}_\mathrm{T}^\mathrm{\,jet} + \vec{p}_\mathrm{T}^\mathrm{\,Z} = 0 \tag{4.2}$$

Assuming that the Z boson is reconstructed correctly, but the jet reconstruction alters its momentum by a response factor $\mathcal{R}$, the sum returns $\vec{p}_\mathrm{T}^\mathrm{\,miss}$.

$$\mathcal{R}\vec{p}_\mathrm{T}^\mathrm{\,jet} + \vec{p}_\mathrm{T}^\mathrm{\,Z} = -\vec{p}_\mathrm{T}^\mathrm{\,miss} \tag{4.3}$$

Thus, inserting Equation 4.2 in Equation 4.3 yields the MPF response variable.

$$\mathcal{R}_\mathrm{MPF} = 1 + \frac{\vec{p}_\mathrm{T}^\mathrm{\,miss} \cdot \vec{p}_\mathrm{T}^\mathrm{\,Z}}{\left(\vec{p}_\mathrm{T}^\mathrm{\,Z}\right)^2} \tag{4.4}$$

In contrast to the p_T balance, the MPF response is hardly sensitive to additional jet activity by taking into account the complete detector information. However, any poorly calibrated detector component can bias the MPF results.

In an inclusive Z + jet event topology, there is usually more than one jet present in the event. To avoid ambiguity, the selection of jets is ordered by transverse momenta and the highest-p_T (*leading*) jet is chosen as the jet to be calibrated. To provide a measure of additional jet activity, an observable α is defined as the ratio of the second-highest-p_T (*sub*leading) jet and p_T^Z.

$$\alpha = \frac{p_\mathrm{T}^\mathrm{jet2}}{p_\mathrm{T}^\mathrm{Z}} \tag{4.5}$$

The estimation of the p_T balance or MPF response as a function of α makes it possible to extrapolate their values to $\alpha \rightarrow 0$, representing an exclusive Z + jet event topology. In this limit, both methods are expected to coincide and to yield the same correction factor by their inverse response.

4.2.2 Measurement and simulation

Residual jet energy corrections are extracted from data and simulation samples made available by the CMS collaboration. The following sections provide information about the ones taken into account for the Z ($\rightarrow \mu\mu$) + jets channel.

Table 4.1: Information about the simulated samples used for the calibration. All samples are combined to one simulation which is used in the following.

label	process	fixed order
DY1J	DY + 1 jet	LO
DY2J	DY + 2 jets	LO
DY3J	DY + 3 jets	LO
DY4J	DY + 4 (and more) jets	LO

Data

The data used for calibration belong to the Double Muon dataset, meaning that two muons have been identified simultaneously by the trigger algorithm.

The data taking period in 2016 has been divided into several eras labelled with letters from B to H, reflecting the time dependence of the detector performance. It was found that several eras show compatible behaviour in terms of jet reconstruction, so that they were grouped into three *intervals of validity* (IOV): beginning (BCD), middle (EF) and end of data taking period (GH). The most important impact on the detector performance was caused by dynamic inefficiencies in the tracking system, an issue whose fix defined the transition of the IOV EF to the IOV GH [52]. Thus, the jet energy corrections are derived separately for each IOV.

The events recorded within a certain time window are grouped into *luminosity* sections [53]. If a luminosity section is certified as *good*, meaning that the detector performance meets the necessary requirements, it is used for the analysis. A data certification file provides the information about good luminosity sections.

The certified data collected in 2016 correspond to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$.

Simulation

To obtain the residual corrections, the results from data have to be compared to simulations. These simulations were created using the Pythia8 event generator, as introduced in Section 2.5. The hard process matrix elements were provided by MadGraph5 at LO precision. Parton showering was used to improve the agreement to data. Their merging with the matrix element was provided by the MLM scheme [25]. The simulation of the interaction of particles with the detector was performed with the Geant4 simulation package.

The Z ($\rightarrow \mu\mu$) + jets process is included in the Drell-Yan process [54] in association with jets (DY + jets). Four independent simulated samples with a different number of jets in the final state of the hard process are used and combined: exactly 1 jet, 2 jets, 3 jets and at least 4 jets, respectively. Table 4.1 provides an overview about the simulated samples taken into account.

4.2.3 Event selection

The events that are suitable for the calibration are chosen as follows according to several requirements on the detector-level event topology. The relevant objects for the calibration are the muons, which are used to reconstruct the Z boson, and the leading and subleading jets.

Muon selection

Muon candidates are reconstructed by the PF algorithm. Muon identification and isolation criteria are applied as introduced in Section 3.3.3. The *tight* identification criterion ensures further rejection of misidentified or background muons, the *tight* isolation ensures the rejection of muons originating from heavy-flavour quark decays. Additionally, muons are only taken into account in a kinematic range of pseudorapidity $|\eta^{\mu}| < 2.3$, determined by the tracker coverage, and transverse momentum $p_{\mathrm{T}}^{\mu} > 20\,\mathrm{GeV}$, set above the trigger threshold in data, which requires two muons to be reconstructed as isolated global or tracker muon, one with a transverse momentum $p_{\mathrm{T}}^{\mu} > 17\,\mathrm{GeV}$ and the second one with $p_{\mathrm{T}}^{\mu} > 8\,\mathrm{GeV}$.

Figure 4.2 shows the distributions of the transverse momenta and pseudorapidities of muons and antimuons.

Z boson selection

The four-momenta of the two muons passing the previous selection steps can be added to form a dimuon system, which is called a Z boson candidate if the muons have opposite charges and the resulting dimuon mass $m^{\mu\mu}$ lies within a difference of $20\,\mathrm{GeV}$ to the official Z boson mass $m_{\mathrm{PDG}}^{\mathrm{Z}} = 91.1876\,\mathrm{GeV}$ [8]. If the event contains three muons, the Z boson candidate is chosen as the one whose mass is closest to $m_{\mathrm{PDG}}^{\mathrm{Z}}$. Events containing more than three muons are vetoed.

Figure 4.3 shows the transverse momentum, the mass and the rapidity of the reconstructed Z boson.

Jet selection

Jets are clustered from PF candidates by the anti-k_{t} algorithm with a distance parameter of $R = 0.4$. The charged hadron subtraction (CHS) algorithm revises reconstructed jets for charged hadronic PF candidates not originating from the reconstructed main collision vertex, thereby reducing the pileup contribution.

The *loose* jet identification criterion, as introduced in Section 3.3.4, rejects noise or badly reconstructed jets. As muons are also reconstructed as jets by the clustering algorithm, each jet with an angular distance $\Delta R < 0.3$ to one of the two muons which are used for the Z boson reconstruction is vetoed to avoid double counting.

The JEC, except for residual or later corrections, are applied. They are applied to the missing transverse momentum $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ by replacing those PF candidates that are clustered

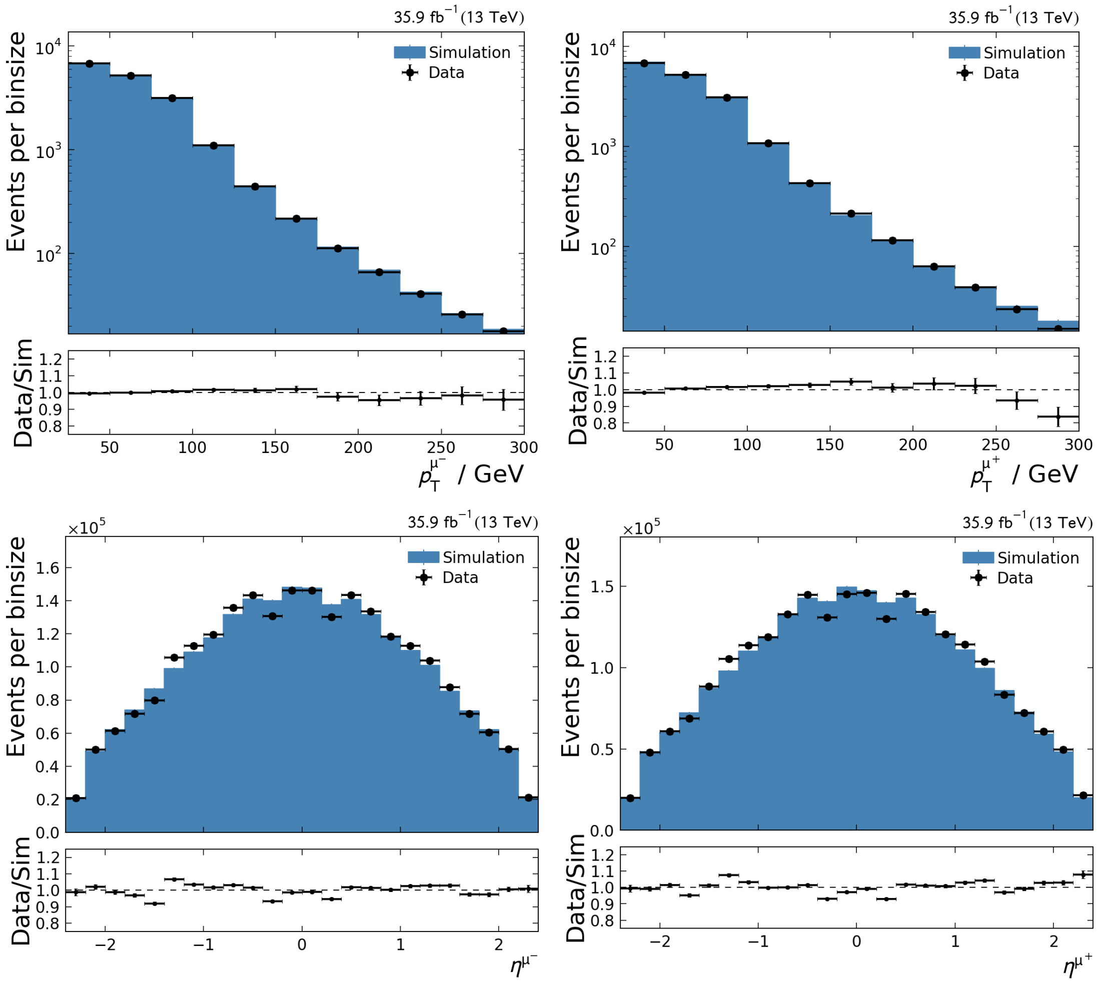


Figure 4.2: Distributions of the transverse momenta (top) and the pseudorapidity (bottom) of the muon (left) and the antimuon (right). The data distributions are normalized to the simulated distributions, resulting in a good agreement. The pseudorapidities of the muons exhibit a dip at certain pseudorapidity values, originating from gaps in the muon system. Muons with tracks in proximity to gaps can be vetoed because of insufficient identification criteria. The ratio hints that these inefficiencies are not completely modelled in simulations. These differences have no impact on the calibration purpose, as only well-reconstructed events are taken into account. For a cross section measurement, corrections have to be applied.

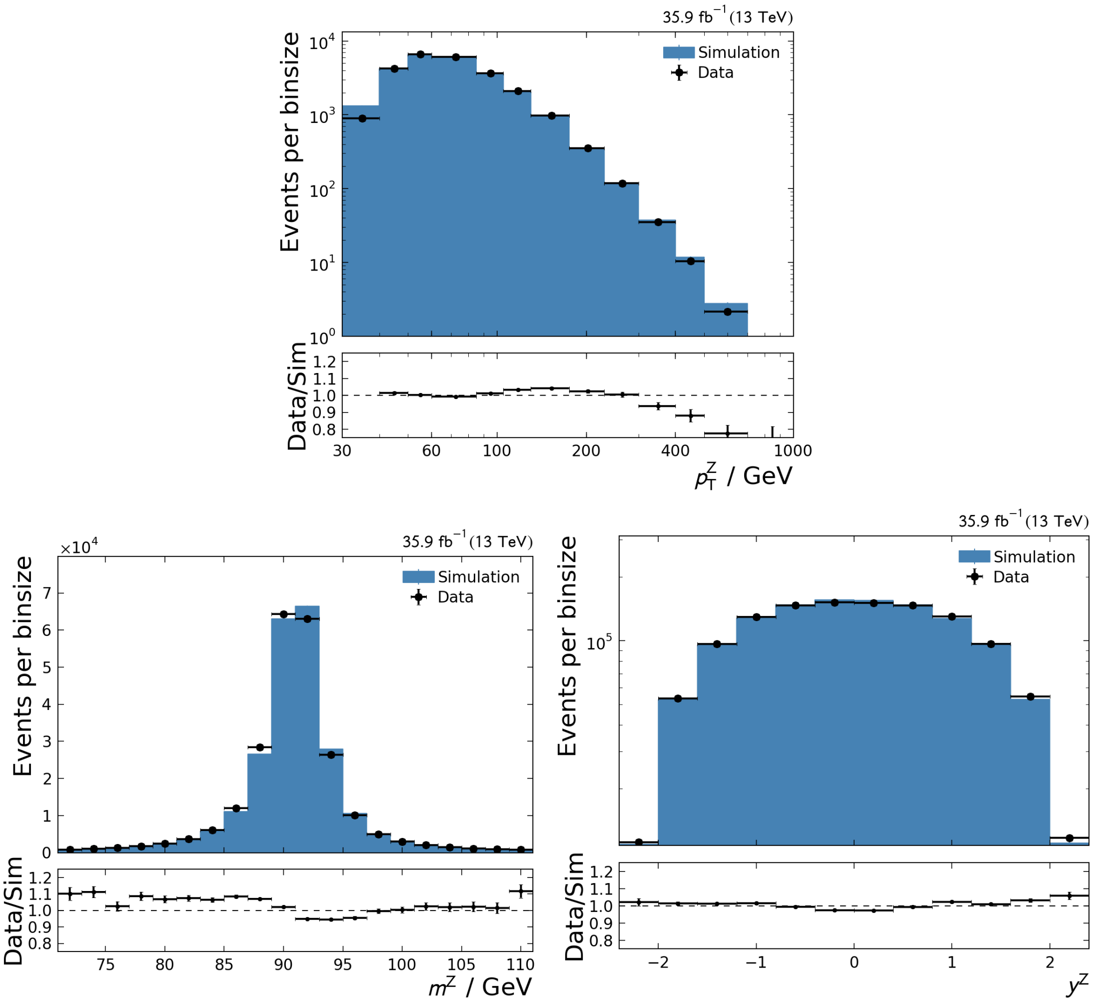


Figure 4.3: Distributions of the transverse momentum (top), the mass (bottom left) and the rapidity (bottom right) of the Z boson. The data distributions are normalized to the simulated distributions, resulting in a good agreement in the entire y^{Z} and a wide $p_{\mathrm{T}}^{\mathrm{Z}}$ range. A small shift of less than 0.2 % can be observed in m^{Z}, which does not affect the jet calibration.

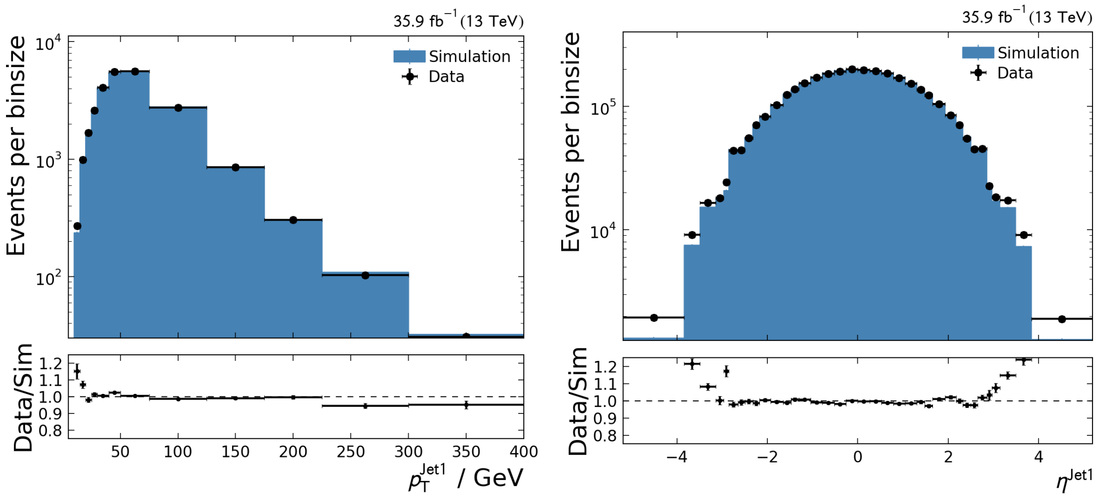


Figure 4.4: Distributions of the transverse momentum (left) and the pseudorapidity (right) of the leading jet. The $|\eta^{\text{jet1}}| < 1.3$ selection was not applied for the right figure. The data distributions are normalized to the simulated distributions, resulting in a good agreement in a wide p_T^{jet1} and the central pseudorapidity region.

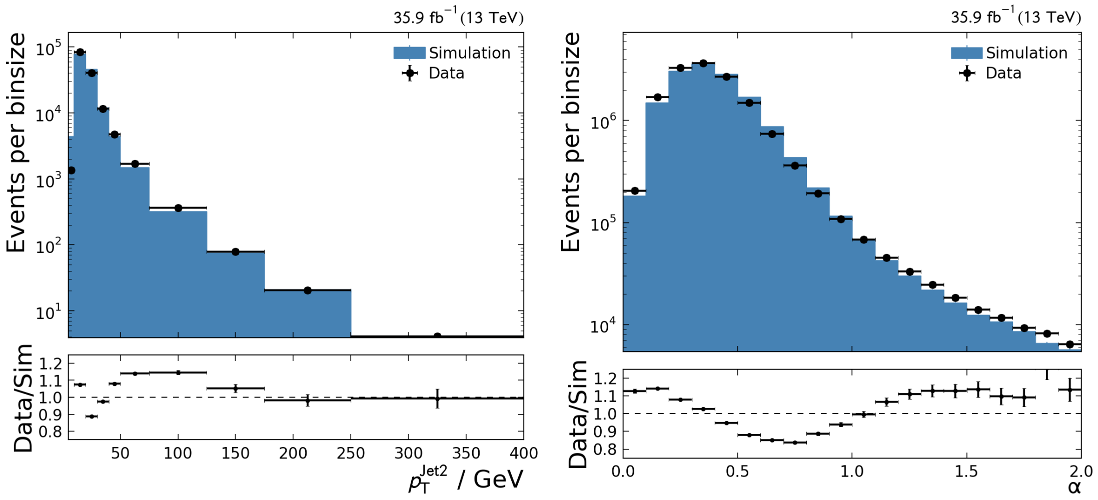


Figure 4.5: Distributions of the transverse momentum of the subleading jet p_T^{jet2} (left) and $\alpha = p_T^{\text{jet2}}/p_T^{Z}$ (right). The $\alpha < 0.3$ selection was not applied for these figures. The data distributions are normalized to the simulated distributions. Additional jet activity is not completely described by the LO simulation. These deviations are compensated in the final results by the extrapolation $\alpha \rightarrow 0$.

into jets by the corrected jets and recalculating $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. This yields the *Type-I* corrected $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ which is then used to calculate the MPF response.

After JEC, the jets are reordered by p_{T} and the leading and the subleading jets are selected. The leading jet is requested to have a transverse momentum of $p_{\mathrm{T}}^{\mathrm{jet1}} > 12\,\mathrm{GeV}$ and a pseudorapidity of $|\eta^{\mathrm{jet1}}| < 1.3$.

Figure 4.4 shows the distributions of the transverse momentum and the pseudorapidity of the leading jet.

$\boldsymbol{\alpha}$ and $\boldsymbol{\Delta\phi}$ selection

To reduce the influence of additional jet activity in the selected events, the second jet is resctricted to $\alpha = p_{\mathrm{T}}^{\mathrm{jet2}}/p_{\mathrm{T}}^{\mathrm{Z}} < 0.3$. Furthermore, the leading jet and the Z boson are requested to exhibit a back-to-back topology, requesting the angle between their transverse momenta to fulfil $|\Delta\phi^{\ \mathrm{jet1,Z}} - \pi| < 0.34$.

Figure 4.5 shows the distributions of the transverse momentum of the subleading jet and the second jet activity α, indicating that additional jet activity is not completely described by the LO simulation. The extrapolation $\alpha \rightarrow 0$ compensates for this sort of data-to-simulation discrepancy.

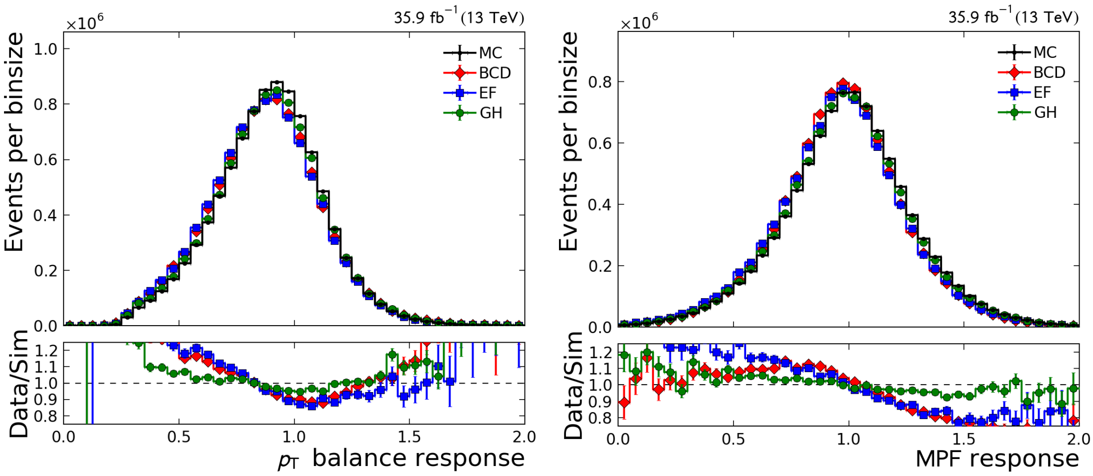


Figure 4.6: Distributions of the p_T balance response (left) and the MPF response (right) separated into the IOVs BCD, EF and GH. The data distributions are normalized to the simulated distributions.

4.2.4 Response estimation

As it was found that the detector performance varies during the data taking period, three IOVs were chosen that for themselves show compatible behaviour in terms of jet reconstruction, labelled with BCD, EF and GH. Figure 4.6 shows the distribution of the two response methods for these IOVs in comparison to the simulation. Whereas the MPF response is distributed around unity, the mean of the p_T balance response is shifted towards lower values as a result of the occurence of additional jets.

A shift towards lower response values can also be observed when comparing the respective distributions from the different IOVs to the ones from the simulation. The aim of residual correction factors is to shift each jet p_T so that the mean values of the data and the simulated response distributions match. Different correction factors are necessary for different IOVs, indicated by varying differences between data and simulation. To derive these factors, the subleading jet activity has to be eliminated.

4.2.5 Extrapolation

The quantity $\alpha = p_\mathrm{T}^\mathrm{jet2}/p_\mathrm{T}^\mathrm{Z}$ was introduced as a measure for additional jet activity. Figure 4.7 shows the mean values of the p_T balance, the MPF and the *true* response as a function of α for the IOV BCD. The true response is defined in simulation only as the ratio of the leading jet p_T on detector level to the one on particle level.

Within the allowed range of $\alpha < 0.3$, the response values exhibit an approximately linear behaviour. To eliminate additional jets and obtain results corresponding to an exclusive Z + jet event topology, the responses are extrapolated to zero using a linear function. The sketch in Figure 4.8 illustrates this procedure. Decreasing $p_\mathrm{T}^\mathrm{jet2}$ forces any

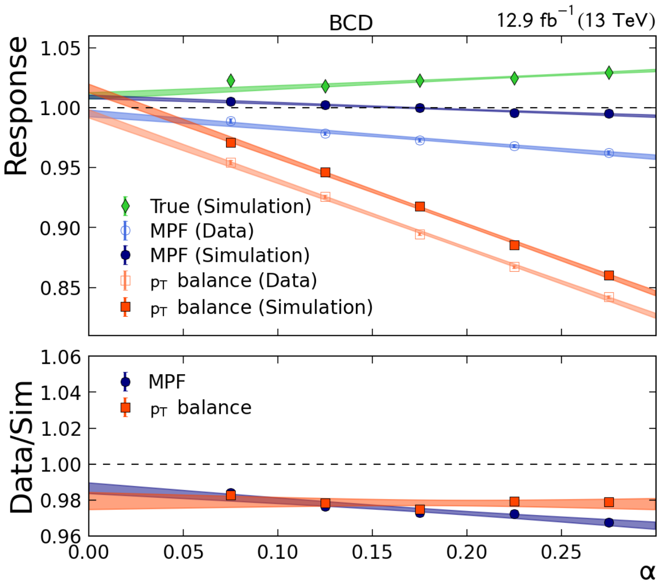


Figure 4.7: Linear extrapolation of the p_T balance, MPF and true response to $\alpha = 0$ for the IOV BCD. The lowest bin $\alpha < 0.05$ is excluded because of the lack of events.

jet except for the leading jet to vanish. The use of α as extrapolation quantity allows for a finer tuning dependent on the p_T^Z region than the use of p_T^{jet2}.

The strong sensitivity of the p_T balance response to the occurence of additional jets is indicated by the slope in Figure 4.7, whereas the MPF response remains stable with respect to α. At $\alpha = 0$, their extrapolated values agree for data and simulation, respectively, within the uncertainties, given as statistical uncertainies propagated through the extrapolation. The agreement provides the verification that the p_T balance and MPF methods coincide in case of an exclusive Z + jet event topology. Furthermore, the agreement of the true response with both p_T balance and MPF response in simulation confirms that the methods yield suitable correction factors for the transverse momentum of a jet.

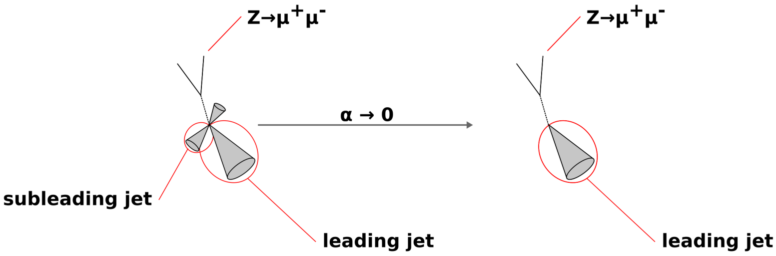


Figure 4.8: Sketch of the extrapolation procedure. Decreasing α forces additional jets to vanish, resulting in an exclusive Z + jet topology.

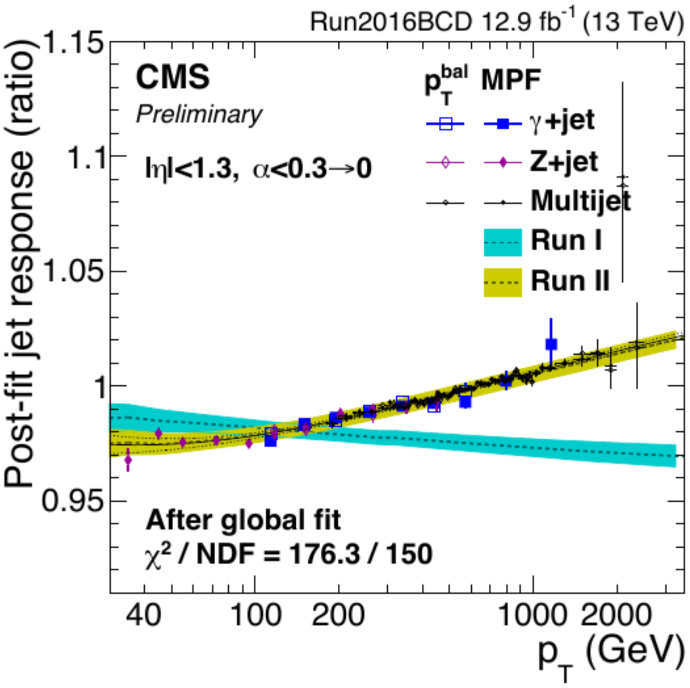


Figure 4.9: Ratio of response values of data to simulation as a function of p_T^{jet} after the global fit (post-fit). The fit result for the IOV BCD is shown as a dashed line with yellow uncertainty bands (Run II). In contrast to the result from the 8 TeV dataset (Run I), which is displayed as the line with blue uncertainty bands, the linear parametrization had to be extended to allow for an adequate fit convergence. Taken from [55].

4.3 Global fit results

The results of the absolute residual correction determination are provided as a function of p_T of the reference object and additional jet activity. The previously explained Z (→ $\mu\mu$) + jet calibration channel is brought together with the Z (→ ee) + jet, which is also provided by our KIT working group, γ + jet and multijet channels that are analysed similarly.

The global fit is used to parametrize the p_T balance and the MPF response from all channels simultaneously as a function of p_T^{jet} after extrapolating to $\alpha \to 0$. Nuisance parameters are taken into account to respect uncertainties from the different channels. The Z (→ $\mu\mu$) + jet channel contributes to these uncertainties by the momentum scale uncertainty for muons of the order of 0.2 %. The result of the global fit for the early data taking period BCD, corresponding to an integrated luminosity of 12.9 fb^{-1}, is presented in Figure 4.9, provided by [55]. The inverse of the fit gives rise to p_T^{jet}-dependent absolute residual correction factors for jets in the low-$|\eta|$ region. In combination with η^{jet}-dependent relative residual correction factors, jet energies can be corrected for differences between data and simulation in the entire phase space which is accessible to the detector.

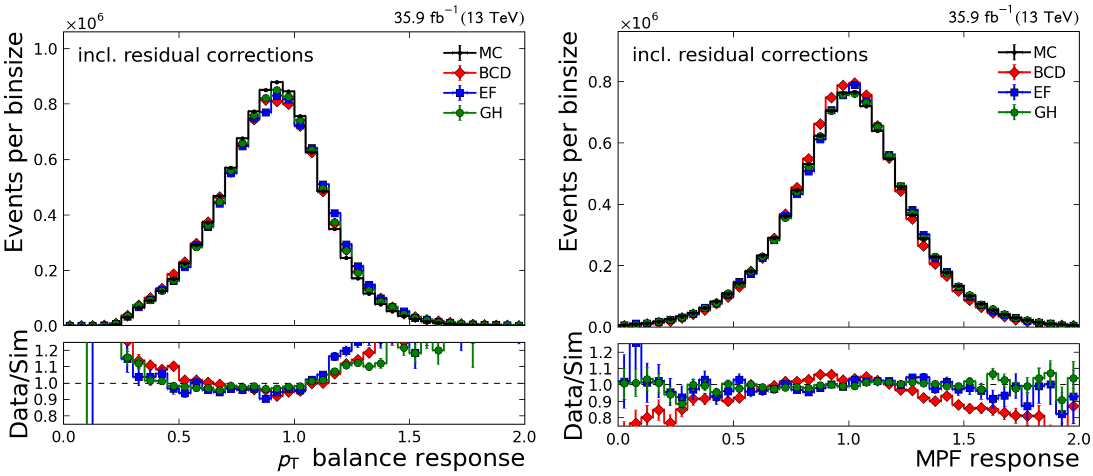


Figure 4.10: Distributions of the p_T balance response (left) and the MPF response (right) separated into the IOVs BCD, EF and GH, after applying the residual corrections. The data distributions are normalized to the simulated distributions.

4.4 Closure test

To verfify the applicability of the methods used to derive both relative and absolute residual corrections, their effect on the p_T balance and the MPF response distributions is presented in Figure 4.10. The ratios indicate that the shift between data and simulation is highly reduced in comparison to Figure 4.6. Especially the MPF response in IOV EF and GH shows a remarkable agreement within their statistical uncertainties.

This impression can be confirmed by Figure 4.11, showing the mean values of the responses as a function of p_T^Z with and without residual corrections applied. It is worth noting that, before applying residual corrections, the response values for the IOV GH steadily lie between the other IOVs and the simulation, which is an artifact of the dynamic inefficiency issue in the tracking system [52]. Due to the residual corrections, the MPF response values from the respective IOVs are consistently brought into agreement with the simulation within 1 % over the entire p_T^Z range. The p_T balance response values are consistently shifted as well, however, they show sizeable differences to the simulation at low p_T^Z.

The illustration of the responses as a function of p_T^Z demonstrates the effect of absolute residual corrections. Figure 4.12 illustrates the effect of relative residual corrections by comparing the mean values of the responses as a function of $|\eta^{\text{jet1}}|$, with and without residual corrections applied. In contrast to the MPF response, the p_T balance response exhibits a significant decrease toward higher $|\eta^{\text{jet1}}|$.

In the pseudorapidity region beyond $|\eta^{\text{jet1}}| < 2.4$, which is not covered by the tracking system, the agreement between data and simulation becomes unstable. Nevertheless, the overall agreement in this region is still improved by the residual corrections.

The differences between p_T balance and MPF response method are included in the JEC uncertainties, as described in Section 4.5.

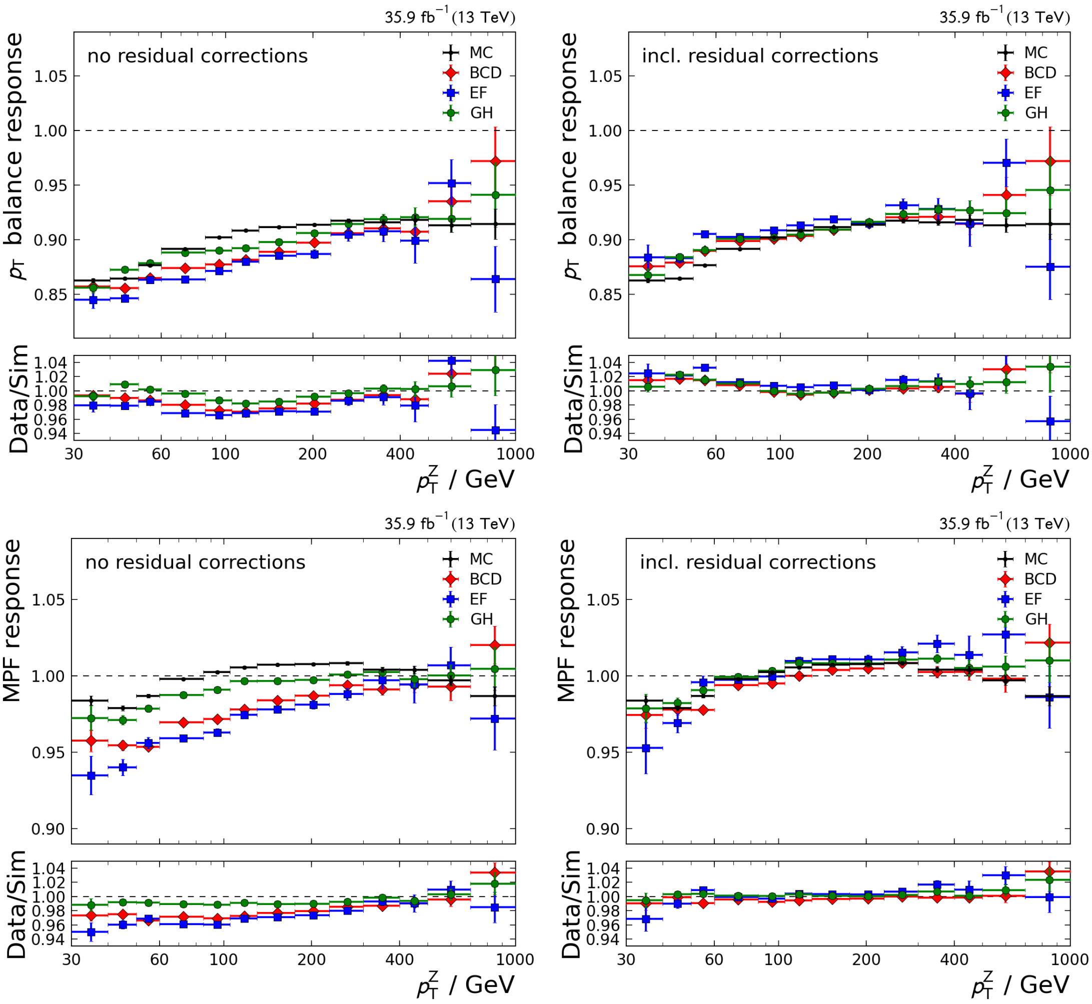


Figure 4.11: The p_T balance response (top) and the MPF response (bottom) as a function of p_T^Z separated into the IOVs BCD, EF and GH, before (left) and after (right) applying residual corrections, compared to the simulation (MC).

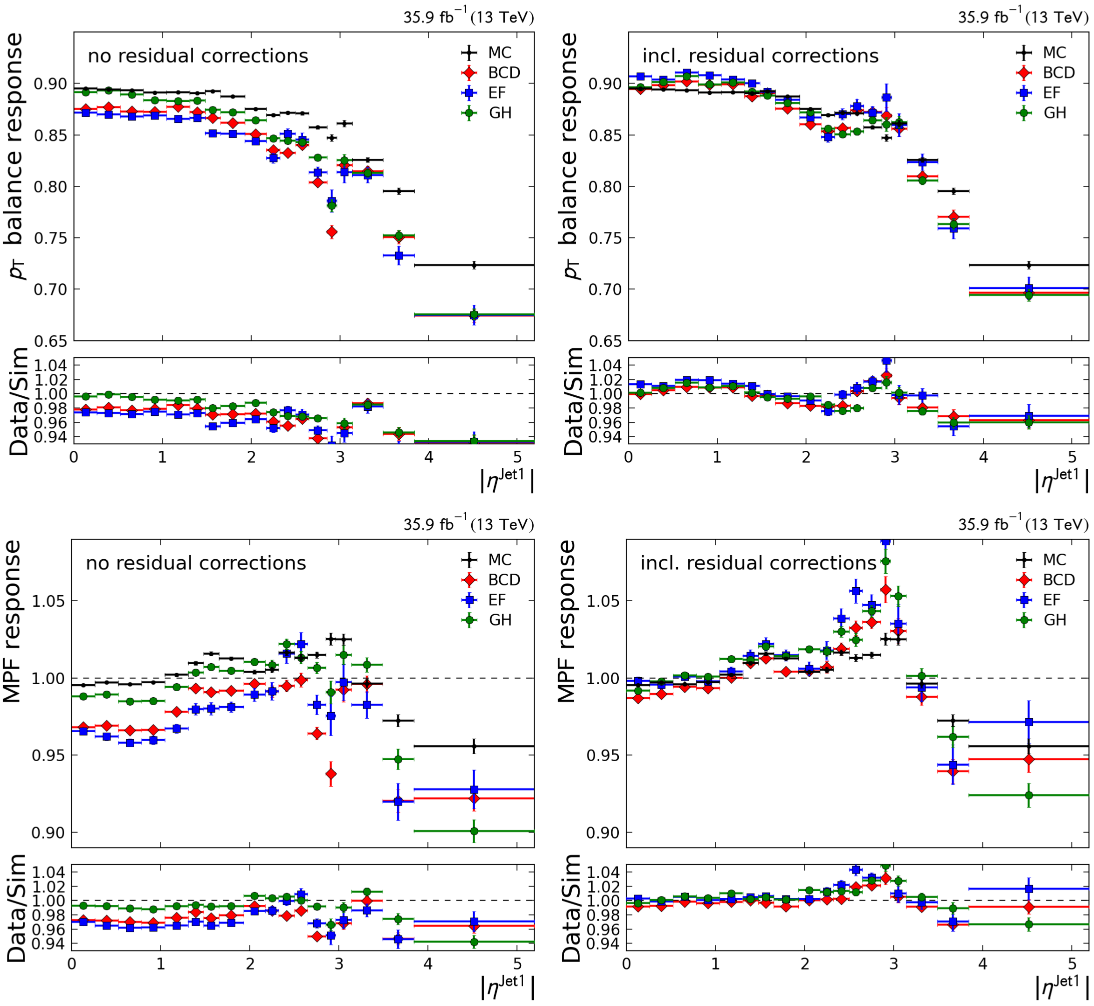


Figure 4.12: The p_T balance response (top) and the MPF response (bottom) as a function of η^{jet1} separated into the IOVs BCD, EF and GH, before (left) and after (right) applying residual corrections, compared to the simulation (MC). The $|\eta^{\mathrm{jet1}}| < 1.3$ selection was not applied in these figures.

4.5 Jet energy correction uncertainties

The results of the full calibration chain are centrally provided for use by analysts of the CMS collaboration. As any measurement, the calibration of jet energies can be biased by systematic effects due to approximations and limited knowledge of detector components. Uncertainties are attached to the derived correction factors to take into account the following effects.

- **Pileup:**

  An uncertainty on offset corrections can be derived from the difference between the true and the measured offset contribution in a simulated event and the variation of the mean number of pileup interactions from an average of $\langle\mu\rangle = 25$. A significant contribution arises at low $p_{\mathrm{T}}^{\mathrm{jet}}$.

- **Relative scale:**

  An uncertainty on both simulated and residual η^{jet}-dependent relative corrections is derived by studying the differences in the physics modelling (parton shower, hadronization and multiple parton interaction) among event generators and in the jet energy resolution between data and simulation.

- **Absolute scale:**

  An uncertainty on both simulated and residual $p_{\mathrm{T}}^{\mathrm{jet}}$-dependent absolute corrections is derived from reconstruction uncertainties of the reference objects (photon, electron and muon) and the differences in the physics modelling among simulations.

- **Jet flavour:**

  An uncertainty on flavour corrections is derived by studying differences among event generators in the simulation of partons with a particular flavour.

- **Time stability:**

  An uncertainty due to varying detector conditions can be derived by comparing the corrections obtained for the full data taking period to the luminosity-weighted results for each era.

- **Method & sample:**

  An additional uncertainty is derived from the difference between the results of the p_{T} balance and the MPF method and among results from different data samples. Thereby remaining discrepancies beyond statistical and systematic uncertainties can be covered.

Figure 4.13 shows the uncertainty results obtained for the 2016 data taking period taken from [55]. The individual uncertainty sources are combined as the square root of their quadratic sum. With an uncertainty of less than 1 %, the highest precision can be achieved for a central pseudorapidity jet with a sufficiently large transverse momentum

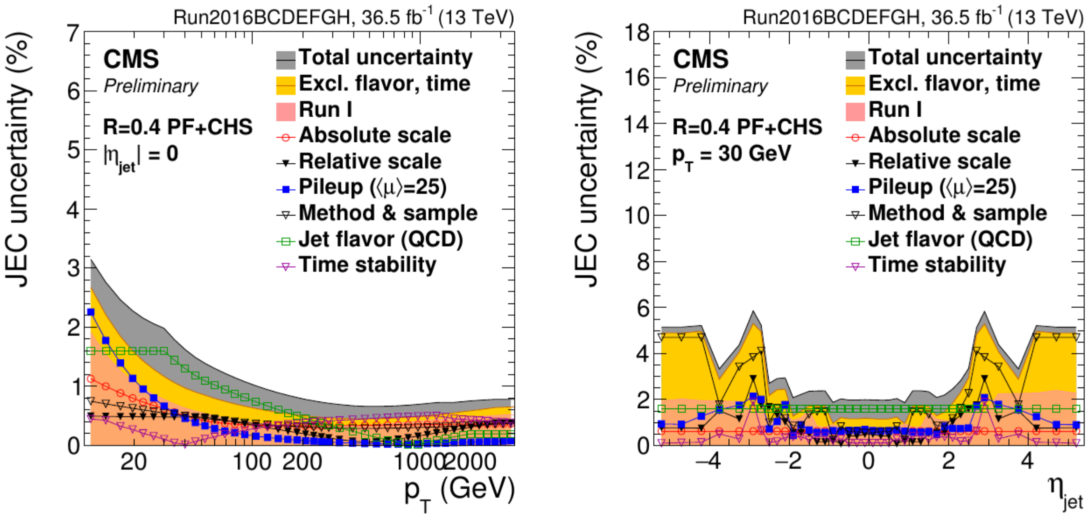


Figure 4.13: Relative jet energy uncertainties as a function of p_T^{jet} in the low η^{jet} region (left) and as a function of η^{jet} in the low p_T^{jet} region (right). The respective uncertainty sources are shown as symbols. The yellow shape represents the total uncertainty when taking into account pileup, relative and absolute scale and method & sample uncertainties. Including time stability and jet flavour uncertainties results in the grey shape. For comparison, the result from the 8 TeV dataset (Run I), excluding time stability and flavour uncertainties, is displayed as red shape. Taken from [55].

of more than 100 GeV. This uncertainty is slightly larger than the one obtained in the 8 TeV dataset (Run I), which is mainly because of the method & sample uncertainty. However, given the challenges with which the collaboration was faced during the data taking at high luminosities, such as increased pileup from an average $\langle\mu\rangle \approx 20$ to $\langle\mu\rangle \approx 25$ and dynamic inefficiencies in the tracking system [52], this value is an indication for the outstanding performance of jet reconstruction in the CMS detector.

4.6 Summary

In this chapter, the workflow of the jet energy calibration within the CMS collaboration was presented. The described methods have been developed for the data taking periods with a collision energy of 7 TeV [50] and 8 TeV [51]. The results of the different calibration steps were provided by several working groups. For the data recorded in the year 2016, with a collision energy of 13 TeV, the analysis of Z $(\rightarrow \mu\mu)$ + jet events used for the determination of absolute residual corrections has been performed and presented within this thesis. The results of a global fit, which combines the results of all channels for the residual corrections to achieve the highest accuracy, have been validated by closure tests. The estimation of the final JEC uncertainties was described.

During the data taking period in 2016, the collaboration faced challenges such as an increase of pileup in comparison to previous periods and a dynamic inefficiency in the tracking system. Especially the fixing of the latter prompted the separation of the data taking period into several IOVs, for which data-driven residual corrections were derived independently. This practice of subdividing data into IOVs in order to better evaluate time dependent effects is maintained for subsequent data taking periods in the years 2017 and 2018.

The calibration results presented within this thesis have been used by various analyses which include the 2016 dataset. However, during the processing of the 2017 and the 2018 calibration, additional detector issues have been discovered that necessitate further studies on their effects on the 2016 jet energy calibration. Because of those ongoing studies the final correction results could not be presented here. The publication of final results on the complete 13 TeV dataset (RunII) is expected within the next years but is beyond the scope of this thesis.

CHAPTER 5

Measurement of the triple differential inclusive Z ($\rightarrow \mu\mu$) + jet cross sections

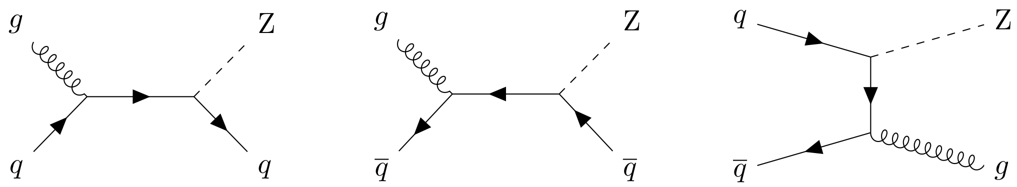
The measurement of cross sections is a central task in particle physics, as it immediately allows to compare experimental results to theoretical calculations. At the LHC, protons collide at energies at which their inner structure is observable. This inner structure can be described by parton distribution functions (PDF). The precision of theoretical calculations is determined by the precise knowledge of the PDFs. Since these functions cannot yet be calculated from first principles, the measurement of a well-understood process can in turn be used for constraints.

The aim of this thesis is to provide inclusive cross section results for Z boson production at a centre-of-mass energy of 13 TeV in association with one jet, where the Z boson decays to a muon-antimuon pair. The measurement of this process has several advantages from an experimental and theoretical point of view which are discussed in Section 5.1. It is compared to theoretical predictions calculated at NNLO precision. The results can subsequently be used to improve the description of the proton PDFs.

A number of analyses within the CMS collaboration addressing a similar task have been published in the previous years, such as measurements of the differential Z boson production cross section at 13 TeV [56] and the differential Z boson production cross section in association with jets at 8 TeV [57] and 13 TeV [58]. A double differential inclusive Z boson analysis using the decay channel into an electron-positron pair at 8 TeV [59] and another one using the decay channel into a muon-antimuon pair at 13 TeV [60] are worth to be mentioned as precursor for this thesis.

A triple differential approach was first used by a dijet measurement at 8 TeV [61], introducing the rapidity variables y^* and y_b, with y^* being half the rapidity separation of the two selected jets and y_b being the boost of their centre-of-mass system. While y^* is given by the scattering angle and therefore dependent on the matrix element of the subprocess, y_b is determined by the momentum fractions of the involved partons. For this reason, cross section contributions originating from the PDFs and those resulting from the matrix element can be disentangled.

Within this thesis, the triple differential approach is first applied to Z ($\rightarrow \mu\mu$) + jet events. The rapidity variables are derived from the Z boson and the leading-p_T jet. The cross section is measured as a function of the Z boson transverse momentum p_T^Z and the



(a) quark-gluon **(b)** gluon-antiquark **(c)** quark-antiquark

Figure 5.1: Feynman diagrams of all parton-parton scattering processes contributing to Z + jet production at LO.

rapidity variables y^* and y_b.

Furthermore, the use of a variable ϕ^*_η, derived from muon angular observables, as an alternative to p_T^Z, enables a cross section measurement with reduced uncertainty because of the better resolution in this variable compared to momentum variables.

In the following sections, the analysis is explained in detail. At first, the advantages of the triple differential inclusive Z ($\rightarrow \mu\mu$) + jet cross section measurement and the used observables are introduced in Sections 5.1 and 5.2. Subsequently, the datasets and simulations which are taken into account are presented in Section 5.3. The details of the object reconstruction are explained in Section 5.4. Unfolding is used to correct the measured distributions for detector effects based on a so-called *forward smearing* aproach, which is described in Section 5.5. Experimental uncertainties associated with the measurements are explained in Section 5.6. The measured cross sections are compared to the simulations in Section 5.7. Theoretical predictions have been computed at NNLO precision as described in Section 5.8. At last, the measured cross sections are compared to the predictions in Section 5.9.

5.1 Z boson production in association with at least one jet at the LHC

As previously discussed in Chapter 4, the analysis of Z boson events in association with jets is of interest especially for tasks that require high experimental precision, such as jet energy calibration. As the CMS detector is most accurate at the reconstruction of muons and antimuons, the muonic decay channel of Z bosons provides a very clear signature with low background contribution.

Even more important, the measurement of the inclusive cross section for Z boson production in association with one jet in proton-proton collisions inherits unique advantages from the theoretical perspective. Figure 5.1 shows the Feynman diagrams of those parton-parton scattering subprocesses that contribute to the cross section at LO and which resemble the Drell-Yan process [54] in association with jets. The contribution of a particular subprocess to the cross section highly depends on the momentum fractions of the partons and is therefore determined by the PDFs.

From a comparison of the subprocesses and the knowledge of PDFs determined from

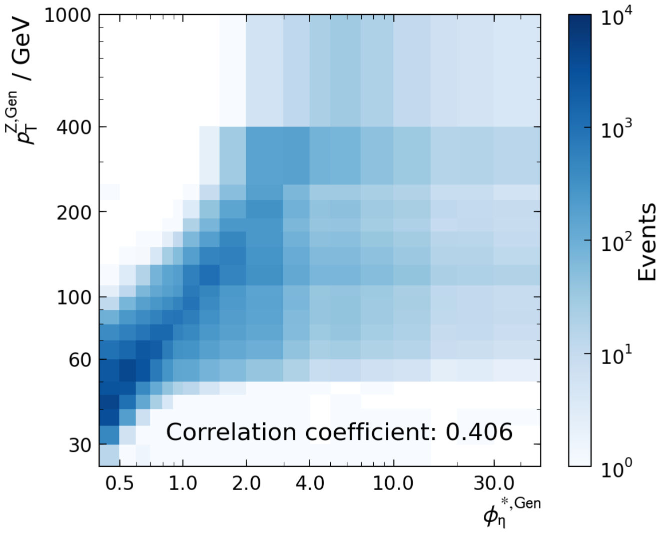


Figure 5.2: 2D histogram of the (ϕ^*_η, p_T^Z) bin occupation filled from an LO simulation to illustrate their correlation.

previous experiments, as already presented in Figure 2.5, a number of conclusions can be made. There is a high probability for a quark to carry a large momentum fraction as well as for a gluon or an antiquark to carry a small momentum fraction. Thus, contributions from the gluon-antiquark subprocess tend to result in events with low energy content. As the PDF of the gluon is much larger than the one of the antiquark, the contribution of the quark-antiquark subprocess is less favoured than the quark-gluon subprocess. This leaves quark-gluon scattering to be expected as the dominant subprocess for the production of inclusive Z + jet events.

Further subprocesses, namely quark-quark, antiquark-antiquark and gluon-gluon scattering, can contribute at NLO or NNLO. As well as the previously discussed subprocesses, the amount of their contribution is highly dependent on the momentum fractions carried by the partons. To gain insight in the subprocess decomposition, experimental observables are required that provide information about these fractions in the reconstructed event. This can be achieved by the rapidity observbles y^* and y_b. The observables are explained in the following section.

5.2 Observables

In an exclusive Z + jet event topology, the momenta of the Z boson and the jet cancel out each other in the transversal plane. Thus, their transverse momenta must be equal. However, due to experimental (e.g. pileup) and theoretical reasons (e.g. gluon radiation), the appearance of additional jets in an event is very likely. To reduce bias by additional jet activity, p_T^Z as the observable with higher experimental reconstruction precision is

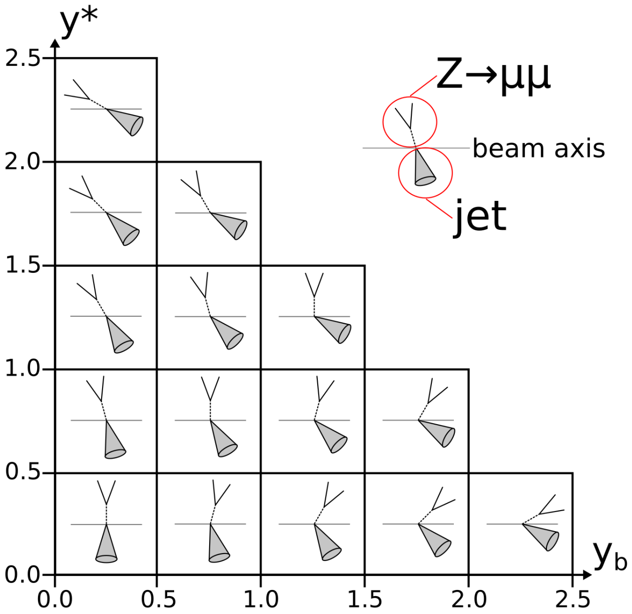


Figure 5.3: Visualization of y^* and y_b. The cone illustrates a jet, the lines illustrate a Z boson that decays to muons, the horizontal line illustrates the beam axis. Large values of y_b correspond to a boosted Z + jet system, large values of y^* correspond to large scattering angles.

preferred over p_T^{jet} or an averaged p_T observable.

For dilepton events, a variable ϕ_η^* can be used to further improve the experimental resolution in comparison to transverse momenta. ϕ_η^* is determined only by angular observables of the muons, namely their pseudorapidities $\eta^{\mu^\pm}$ and their azimuthal angles $\phi^{\mu^\pm}$.

$$\phi_\eta^* = \tan\left(\frac{\pi}{2} - \frac{\phi^{\mu^+} - \phi^{\mu^-}}{2}\right)\sin\left(\theta^*\right) \tag{5.1}$$

$$\cos\left(\theta^*\right) = \tanh\left(\frac{\eta^{\mu^+} - \eta^{\mu^-}}{2}\right) \tag{5.2}$$

It can be thought of as the transverse momentum of the dilepton system divided by the dilepton invariant mass, thus conserving the p_T^Z dependence, but cancelling out the effects of momentum resolution [62]. Since angular observables can be measured more precisely than energy-related ones, ϕ_η^* is expected to be measured with a better resolution than p_T^Z. On the other hand, as can be seen in Figure 5.2, the observables are correlated, such that p_T^Z and ϕ_η^* can be treated analogously.

The rapidities y^Z and y^{jet} are the variables that give an indication to the initial partonic states of the event. Instead of their immediate utilization, it is convenient to combine

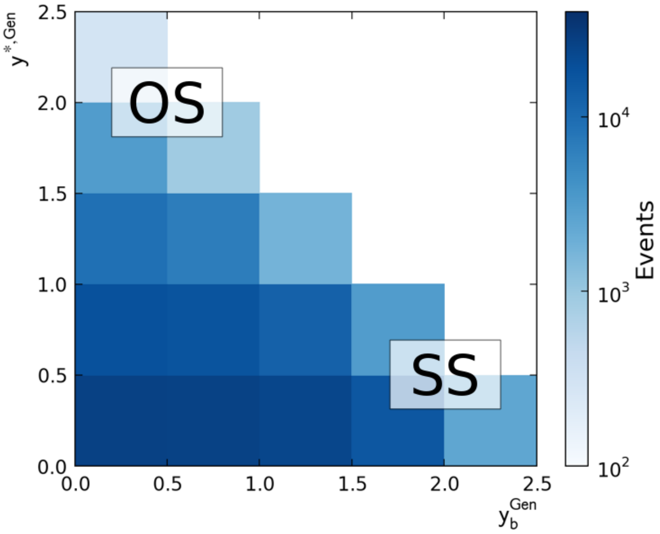


Figure 5.4: 2D histogram of the (y^*, y_b) bin occupation filled from an LO simulation to illustrate the asymmetry between the distribution of same-side (SS) and opposite-side (OS) events.

them to the observables y^* and y_b. This approach was necessary for the dijet analysis [61] to avoid ambiguities of jets as indistingishable objects. As the Z boson is distinguishable from a jet, the ambiguity is no longer present in a Z + jet analysis. However, as explained in the following, the advantage of sensitivity to the PDFs remains.

The observable y^* is defined as the rapidity *separation*

$$y^* = \frac{1}{2}|y^\mathrm{Z} - y^\mathrm{jet}| \tag{5.3}$$

It is closely related to the scattering angle in the centre-of-mass system, which itself is determined by the matrix element of the subprocess. In turn, information about the partonic momentum fractions x_1 and x_2 is contained within the observable y_b, defined as the *boost* of the centre-of-mass system.

$$y_\mathrm{b} = \frac{1}{2}|y^\mathrm{Z} + y^\mathrm{jet}| \tag{5.4}$$

Partons carrying similar values of x_1 and x_2 lead to a small boost, otherwise to large boost.

This results in a natural way to distinguish opposite-side (OS) from same-side (SS) events, where both objects either lie in opposite or the same rapidity directions. A geometric visualization of possible Z boson and jet rapidity configurations in the respective y^* and y_b regions can be found in the sketch in Figure 5.3.

Figure 5.4 illustrates the event distribution in the y^*-y_b-plane, showing an asymmetry between the number of OS and SS events. This asymmetry can be explained by differences in the parton-parton scattering subprocess decomposition in the respective phase space regions. As pointed out in Section 5.1, inclusive Z + jet production is expected to be dominated by the quark-gluon subprocess.

An event in the high y^* region requires both partons to carry similar momentum fractions, which is unlikely for the quark-gluon subprocess, as well as for the quark-antiquark subprocess. Therefore, a proper description of the OS region requires calculations at NLO or NNLO precision that include further subprocesses, such that large NLO k-factors can be expected.

An event in the high y_b region results from partons with different momentum fractions. The quark-gluon subprocess can therefore be expected to dominate this phase space region almost exclusively. This region will be of particular interest for gluon PDF studies.

A three-dimensional phase space spanned by the variables $(p_\mathrm{T}^\mathrm{Z}, y^*, y_\mathrm{b})$ and $(\phi_\eta^*, y^*, y_\mathrm{b})$ permits access to the kinematics of the colliding partons, with the latter providing an even higher precision. The differential cross section measurement based on this approach can subsequently be used for PDF studies.

5.3 Measurement and simulation

The data as well as simulation samples are made available by the CMS collaboration. The following sections provide information about the ones taken into account for the inclusive Z ($\rightarrow \mu\mu$) + jet cross section analysis.

5.3.1 Data

Datasets contain events that are classified according to certain trigger requirements. The data used within this analysis belong to the Single Muon dataset, meaning that at least one muon has been identified by the trigger algorithm.

As already described in Section 4.2.2, the data taking period in 2016 has been divided into several intervals of validity (IOV), reflecting the time dependence of the detector performance. The certified data collected in 2016 correspond to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$.

5.3.2 Simulations

To gain a deeper understanding of the process of interest, simulations were created and will be compared to the measurement. The event generators PYTHIA8 and HERWIG++ were used to generate events, as introduced in Section 2.5. For complicated processes, the matrix elements were calculated by matrix element generators such as MADGRAPH5_aMC@NLO. The simulation of parton showers was used to improve the description of the process. The subsequently used hadronization method depends on the event generator. A simulation of the interaction of particles with the detector was performed with the GEANT4 simulation package.

The inclusive Z ($\rightarrow \mu\mu$) + jet process is described by a Drell-Yan process [54] in association with jets. Several simulated samples have been used:

Table 5.1: The simulated samples used for this analysis. The same labels are also used in subsequent figures.

label	process	fixed order
Signal		
DY (P8+MG)	DY + jets	LO
DY (HW+MG)	DY + jets	LO
DY (P8+aMC)	DY + jets	NLO
Background		
ZZ	Z + Z	LO
WZ	W + Z to leptons + jets	NLO
WW	W + W to leptons	NLO
TT	top quark-antiquark pair + jets	LO
TW	single top quark/antiquark + W	NLO

- **P8+MG**: A simulation at LO precision created with Pythia8 using MadGraph5 and the MLM parton shower merging scheme is used as reference simulation for subsequent figures. To improve the normalization to the luminosity of the data, its total cross section is scaled by a NNLO k-factor.

- **HW+MG**: To study the difference between hadronization models, a simulation at LO precision created with Herwig++ using MadGraph5 and the MLM parton shower merging scheme is taken into account.

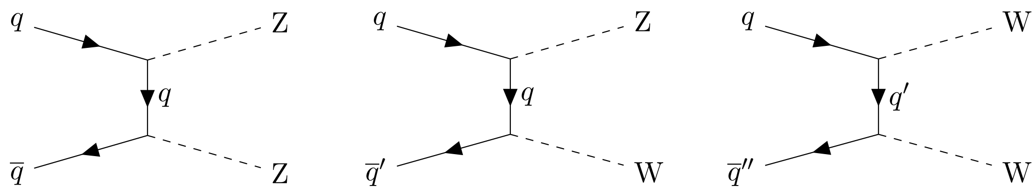
- **P8+aMC**: For a comparison to NLO precision, a simulation created with Pythia8 using MadGraph5_aMC@NLO and the FxFx multijet merging scheme is taken into account.

Apart from the Drell-Yan process, processes whose final states can be misidentified as inclusive Z ($\rightarrow \mu\mu$) + jet events have to be included. The contribution of these background processes needs to be estimated to correct the measured cross sections. Besides the Z boson decay channels to a tau lepton-antilepton pair (Z $\rightarrow \tau\tau$), which is already included in the simulation of the DY + jets process, there are two sorts of processes likely to be misidentified as inclusive Z ($\rightarrow \mu\mu$) + jet events: diboson production, where muons can originate from any boson, and top quark production, where muons originate from W bosons resulting from a top quark decay to a bottom quark.

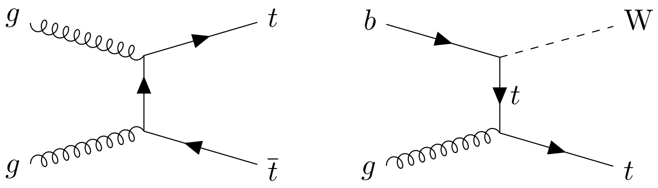
- **ZZ**: The production of two Z bosons was simulated with Pythia8, relying on a built-in LO matrix element calculation.

- **WZ**: The production of a Z boson and a W boson in association with jets, restricted to the leptonic decay channel of both bosons, has been simulated with Pythia8 at NLO precision with MC@NLO parton shower matching.

- **WW**: The production of two W bosons both decaying leptonically has been simulated with Pythia8 at NLO with Powheg parton shower matching.



(a) ZZ **(b)** WZ **(c)** WW



(d) TT **(e)** TW

Figure 5.5: Feynman diagrams of background processes which can be falsely identified as inclusive Z + jet event.

- **TW**: Single top quark or antiquark production in association with a W boson has been simulated with Pythia8 at NLO with Powheg parton shower matching. Two equivalent simulations were combined, one of them representing the top quark and the other one the top antiquark contribution.

- **TT**: The production of top quark-antiquark pairs has been simulated by Pythia8 using a LO MadGraph5 matrix element merged with parton showers according to the MLM scheme.

Table 5.1 provides an overview of the simulated samples taken into account. Feynman diagrams of these processes are presented in Figure 5.5. Some further simulation samples, such as single top quark or antiquark production in the s- and t-channels and W production in association with jets have been examined and yield no significant background contributions.

5.4 Object reconstruction and selection

In the CMS detector, objects are reconstructed using the Particle Flow algorithm which takes into account information from all detector subsystems. The observables which are relevant for the triple differential inclusive Z + jet cross section measurement have been explained in Section 5.2.

To access these observables, the Z boson and one jet have to be reconstructed in an event. The critera regarding the object reconstruction and their selection, which are chosen similarly to Chapter 4, are presented in the following.

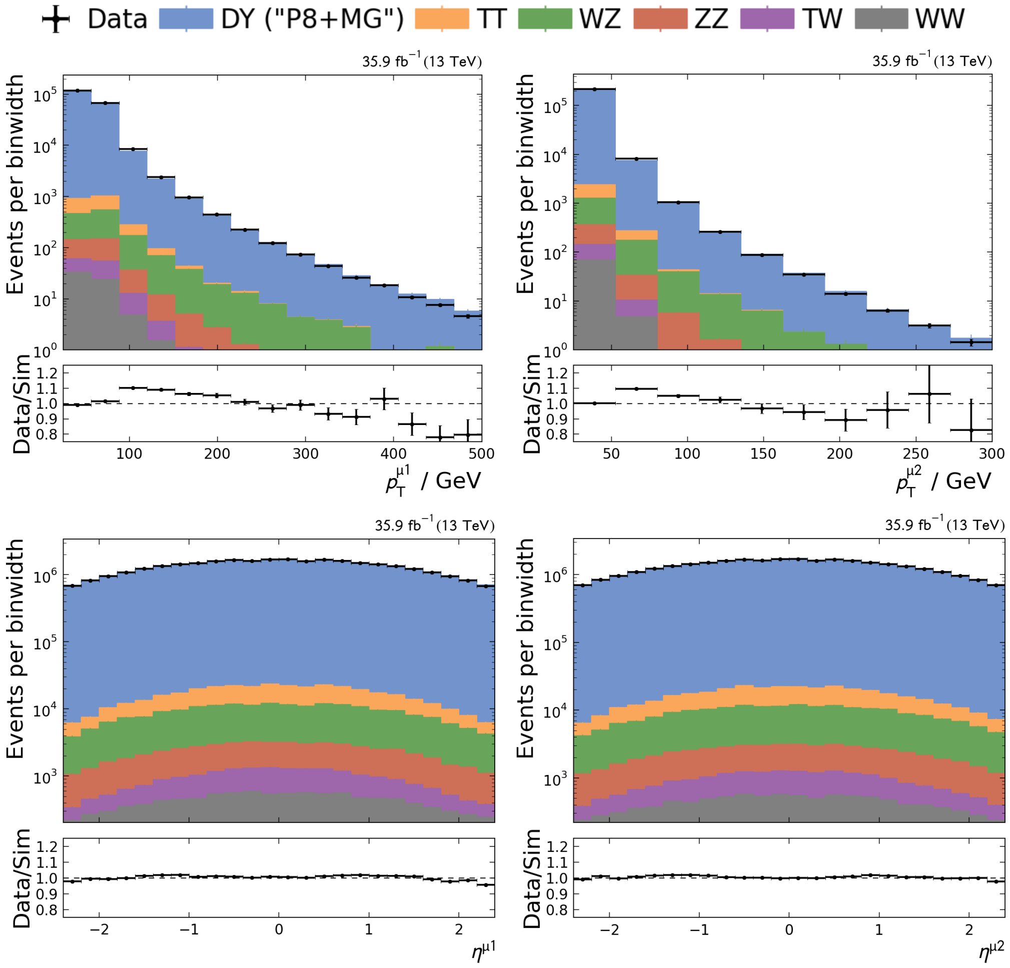


Figure 5.6: Transverse momentum (top) and pseudorapidity distributions (bottom) of the leading-p_T muon (left) and the subleading-p_T muon (right). The simulated distributions are normalized to the luminosity of the data, resulting in good agreement.

5.4.1 Muon selection

Muon candidates are reconstructed by the Particle Flow algorithm. Additional selection criteria ensure further rejection of misidentified or background muons as introduced in Section 3.3.3. The *tight* identification criterion ensures further rejection of misidentified or background muons, the *loose* isolation reduces the number of muons originating from heavy-flavour quark decays.

Corrections are applied on the transverse momentum p_T^{μ} of a muon to correct for geometric properties such as detector misalignment (*Rochester corrections*) [63, 64]. Furthermore, muons are only taken into account in a kinematic range of pseudorapidity $|\eta^{\mu}| < 2.4$ and transverse momentum $p_T^{\mu} > 25\,\text{GeV}$, set slightly above the trigger threshold in data,which requires at least one muon to be reconstructed as an isolated global or tracker muon with a transverse momentum $p_{\mathrm{T}}^{\mu} > 24\,\mathrm{GeV}$.

The identification and isolation criteria are applied equivalently to data and simulation.

A trigger requirement has explicitly been used in the dataset classification. It is therefore applied to data but not necessarily to simulations.

The use of identification and isolation criteria for the selection suppresses the appearance of misidentified muons. Strong requirements, on the other hand, increase the probability of rejecting muons that are suitable for the analysis. The efficiency ε as the fraction of accepted muons for an identification or isolation working point was estimated via tag-and-probe by the CMS collaboration [65]. It was found that the efficiency within Monte Carlo simulations slightly differs from the efficiency in data, as it has been observed in Figure 4.2. This difference is compensated by attaching a weight w to each event following from the Data/MC scale factors $SF(\mu_i)$ for the muon μ_i.

$$SF(\mu_i) = \frac{\varepsilon^{\mathrm{MC}}(\mu_i)}{\varepsilon^{\mathrm{Data}}(\mu_i)} \tag{5.5}$$

These scale factors are derived for both identification and isolation. Since these are independent for both muons in the event, the weight w can be expressed as a multiplicative combination:

$$w = SF_{\mathrm{ID}}(\mu_1)SF_{\mathrm{ID}}(\mu_2)SF_{\mathrm{Iso}}(\mu_1)SF_{\mathrm{Iso}}(\mu_2) \tag{5.6}$$

In a similar way, the efficiencies for the trigger selection in data are compensated by the use of a weight w_{trigger} attached to each event. As it is sufficient for one muon to be triggered, correlations have to be taken into account for the correct efficiency estimation. The weight is therefore given by:

$$w_{\mathrm{trigger}} = \frac{1}{1-(1-\varepsilon_{\mathrm{trigger}}(\mu_1))\,(1-\varepsilon_{\mathrm{trigger}}(\mu_2))} \tag{5.7}$$

Distributions describing the kinematics of the muons are shown in Figure 5.6, showing a good agreement with differences of less than 10 % between data and the simulation.

5.4.2 Z boson selection

The four-momenta of two muons passing the previous selection steps can be added to form a dimuon system, which is called a Z boson candidate if the muons have opposite charges and the resulting dimuon mass $m^{\mu\mu}$ lies within 20 GeV of the official Z boson mass $m_{\mathrm{PDG}}^{Z} = 91.1876\,\mathrm{GeV}$ [8]. If the event contains more than two muons, more than one combination leads to a dimuon system. In this case, the Z boson candidate is chosen as the one whose mass is closer to m_{PDG}^{Z}.

Distributions describing the Z boson kinematics are shown in Figure 5.7. The distributions of $p_{\mathrm{T}}^{\mathrm{Z}}$ and ϕ_{η}^{*} show small deviations of less than 20 % between data and simulation, which is acceptable for an LO simulation.

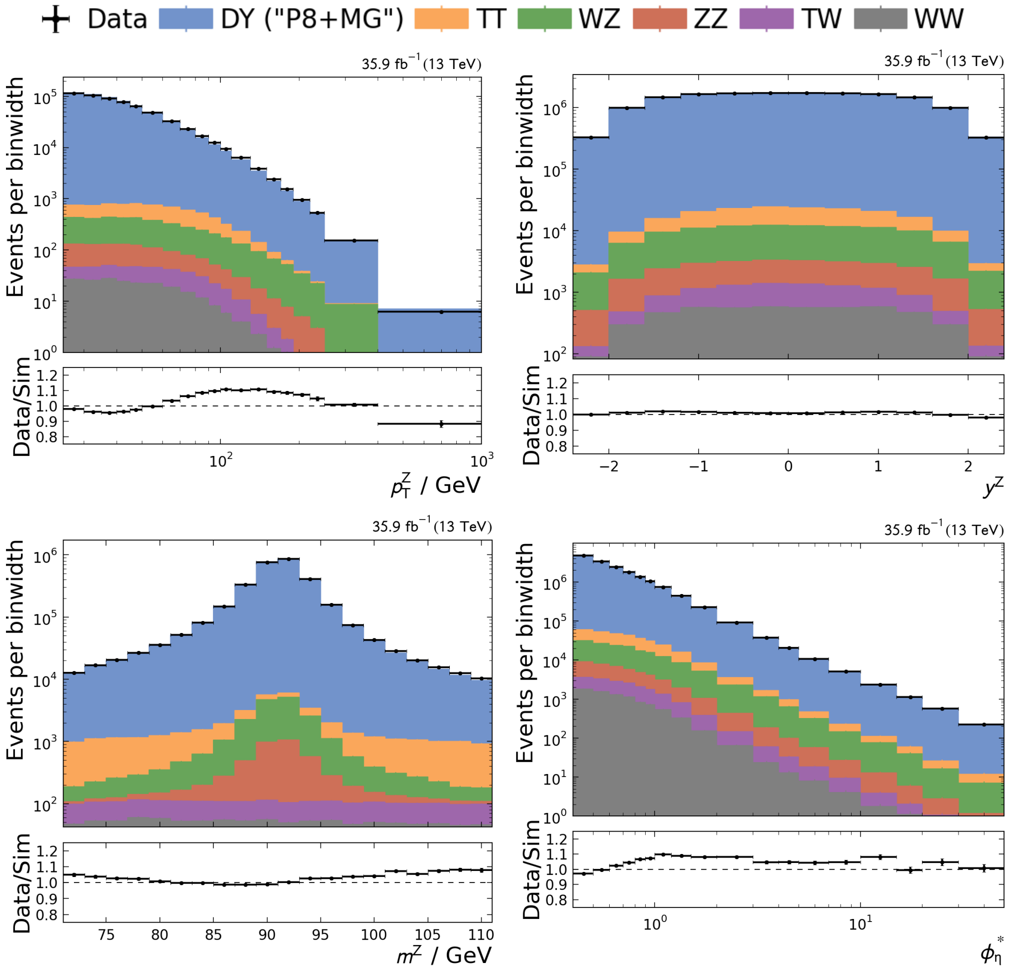


Figure 5.7: Transverse momentum (top left), rapidity (top right) and mass distributions (bottom left) of the Z boson and the ϕ^*_η distribution (bottom right). The simulated distributions are normalized to the luminosity of the data.

5.4.3 Jet selection

Jets are clustered from Particle Flow candidates by the anti-k_t algorithm with a distance parameter of $R = 0.4$, as introduced in Section 2.6. The CHS algorithm revises reconstructed jets for charged hadronic PF candidates not originating from the reconstructed hard scattering vertex, thereby reducing the pileup contribution. To improve its performance, the selection is limited to jets in a rapidity range of $y^{\mathrm{jet}} < 2.4$, where the tracking system is able to identify charged hadrons.

Further selection criteria are applied, as introduced in Section 3.3.4. The *loose* jet identification rejects noise or badly reconstructed jets. The *medium* pileup jet ID working point is used to veto jets likely to originate from pileup.

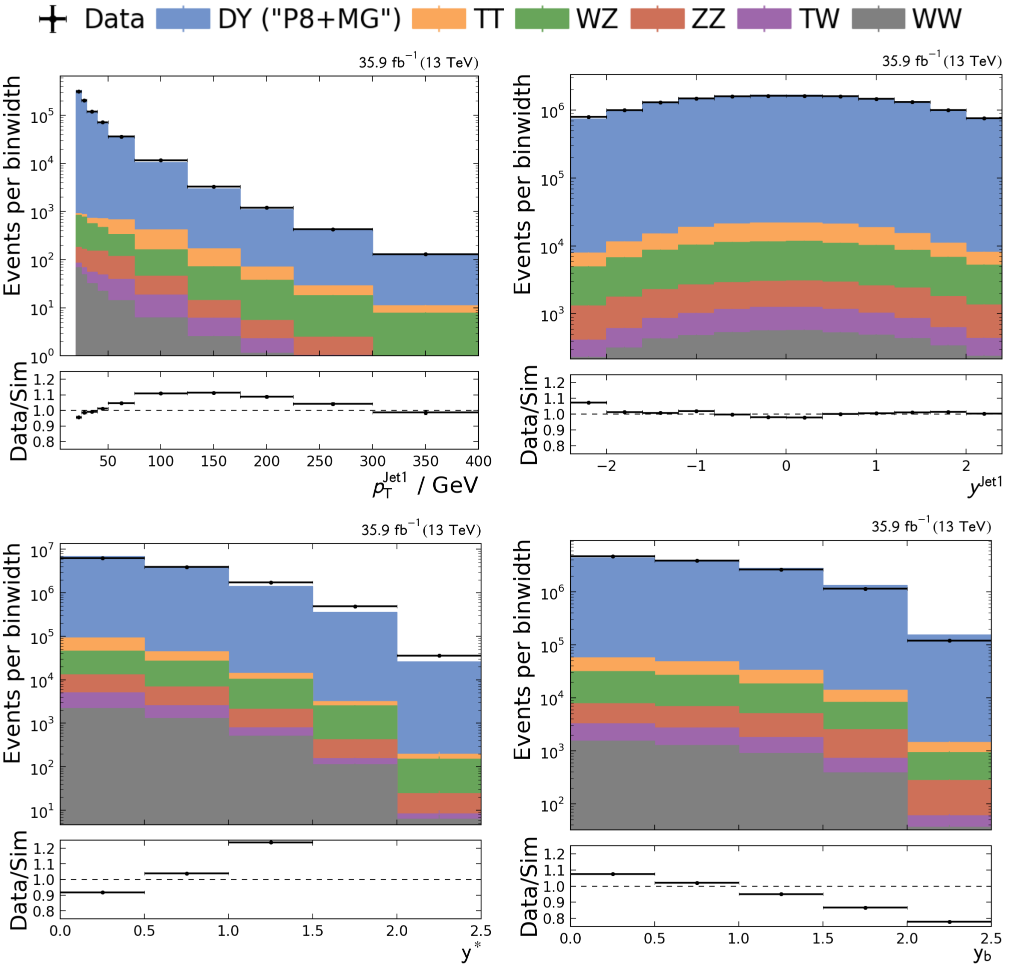


Figure 5.8: Transverse momentum (top left) and rapidity distributions (top right) of the leading jet, y^* (bottom left) and y_b distribution (bottom right). The simulated distributions are normalized to the luminosity of the data.

As even isolated muons are reconstructed as a jet by the clustering algorithm, each jet with an angular distance $\Delta R < 0.3$ to one of the two muons which are used for the Z boson reconstruction is vetoed.

The remaining jets are ordered by p_T after applying the jet energy corrections and the highest-p_T (leading) jet is selected. Each event is further requested to contain a leading jet with a transverse momentum of $p_T^{\text{jet1}} > 20\,\text{GeV}$.

Distributions resulting from the kinematics of the reconstructed leading jet are shown in Figure 5.8. Whereas the distributions of y^{jet1} and y^{Z} are well described by the simulation, y^* and y_b exhibit sizeable deviations. Especially for the description of the y^* distribution, calculations at higher orders are required, as discussed in Section 5.1.

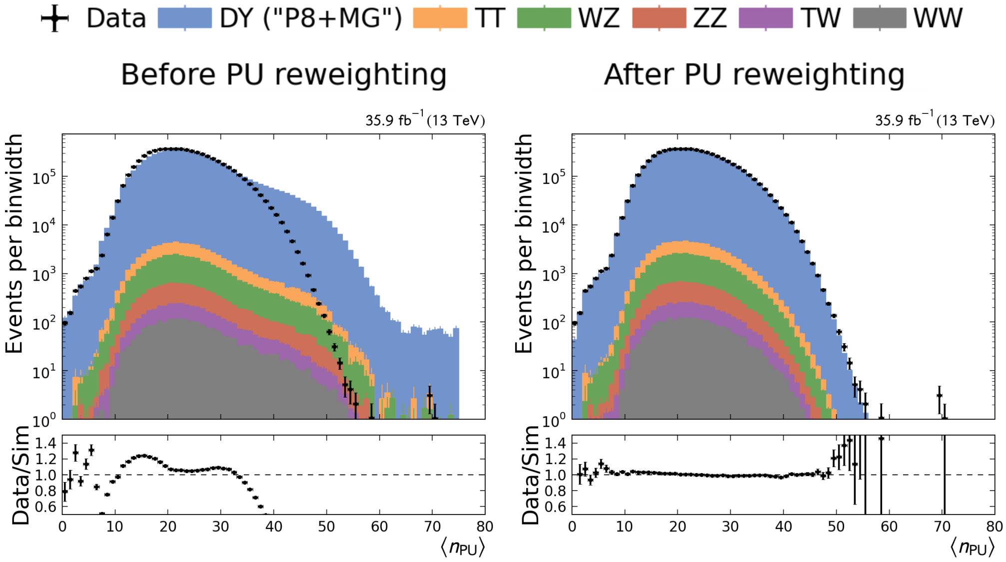


Figure 5.9: Distributions of the average number of pileup events per lumi section $\langle n_{\mathrm{PU}} \rangle$ before (left) and after pileup reweighting (right).

5.4.4 Pileup reweighting

The distribution of the number of pileup interactions n_{PU} in the simulation is only an estimate of the distribution in the experiment. In data, the corresponding accessible quantity is $\langle n_{\mathrm{PU}} \rangle$, meaning the average number of pileup interactions per lumi section, which can be measured from beam properties. To improve the agreement between data and simulation, each simulated event is weighted according to its $\langle n_{\mathrm{PU}} \rangle$ value, such that the distribution of all events matches the actual distribution in data. The effect of pileup reweighting on the $\langle n_{\mathrm{PU}} \rangle$ distribution is shown in the top row of Figure 5.9, leading to a good agreement between data and simulation.

5.4.5 Binning

The measurement of differential cross sections is based on the counting of event numbers by filling histograms, which makes the choice of an adequate histogram binning essential. The binning results from a compromise between the granularity, the statistical precision and the resolution of the observable.

A finite number of bins implies a loss of information, although necessary to ensure an acceptable statistical precision in an experiment with a finite number of events. Even if this number is sufficiently high, a reasonable minimal bin width is determined by the detector resolution. A bin width below the detector resolution gives rise to migrations across bins, causing challenges for unfolding, which is a common procedure used for the correction of detector effects.

For angular variables such as muon and jet rapidities, the resolution is expected to be

Table 5.2: Bin boundaries of triple differential binning in y^*, y_b and p_T^Z or ϕ^*_η.

observable	bin edges
y^*	0.0, 0.5, 1.0, 1.5, 2.0, 2.5
y_b	0.0, 0.5, 1.0, 1.5, 2.0, 2.5
central binning (C)	
$p_\mathrm{T}^\mathrm{Z}/\,\mathrm{GeV}$	25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 130, 150, 170, 190, 220, 250, 400, 1000
ϕ^*_η	0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.2, 1.5, 2.0, 3.0, 4.0, 5.0, 7.0, 10, 15, 20, 30, 50
edge binning (E)	
$p_\mathrm{T}^\mathrm{Z}/\,\mathrm{GeV}$	25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 130, 150, 170, 190, 250, 1000
ϕ^*_η	0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.2, 1.5, 2.0, 3.0, 5.0, 10, 50
extra binning (X)	
$p_\mathrm{T}^\mathrm{Z}/\,\mathrm{GeV}$	25, 30, 40, 50, 70, 90, 110, 150, 250
ϕ^*_η	0.4, 0.6, 0.8, 1.0, 5.0

low in comparison to the energy resolution. The large spacing of the chosen y^* and y_b binning in Table 5.2 ensures that the resolution limit is not reached.

The p_T^Z and the ϕ^*_η binning corresponds to that chosen by previous CMS measurements [56, 58]. A resolution study, which is part of the unfolding Section 5.5.2, asserts that the bin widths remain well above the detector resolution.

However, modifications have been made in certain phase space regions where the event count per bin becomes insufficient to guarantee statistical significance. The different binnings for p_T^Z and ϕ^*_η are listed in Table 5.2. In the central rapidity regions, a central binning (C) is used. An edge binning (E) has to be used in the edge regions, such that the resulting bins contain at least 50 events in the P8+MG simulation sample, which are necessary for further studies that are part of the unfolding Section 5.5. Likewise, the bin containing the maximal y^* values needs to be modified further to fulfil this requirement, resulting in an extra binning (X).

Which p_T^Z or ϕ^*_η binning is applied in which rapidity bin can be retrieved from the map in Figure 5.10. The triangular structure is a result of the phase space choice: no Z boson or jet rapidity observables beyond $|y| = 2.4$ have been taken into account as a consequence of detector properties.

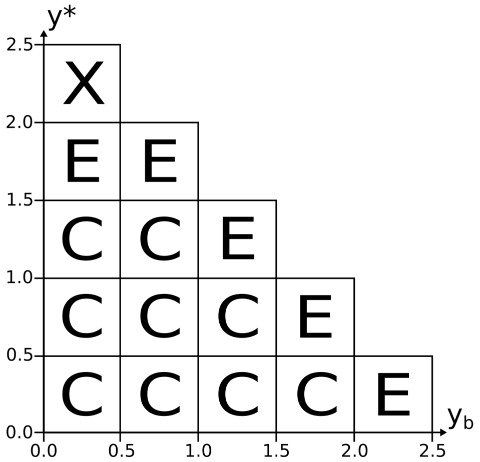


Figure 5.10: Map of which p_T^Z or ϕ_η^* binning, as defined in Table 5.2, is used in which rapidity bin.

5.5 Unfolding

To facilitate a comparison of the measurement to other experimental results or theoretical calculations, the measured distributions have to be corrected for effects that result from the impact of the detector itself. In case of a cross section measurement, the main effects of detector bias are inefficiencies and migrations across measurement bins.

Inefficiencies will occur if an object of interest is for example not reconstructed inside the detector acceptance. Apart from inefficiencies, it is also possible that a different object is falsely identified as an object of interest.

Migrations arise from the fact that a quantity can only be determined with a limited resolution, such that its measured value differs from the original value. Their difference is randomly distributed, such that the original value can not be recovered on an event-by-event basis.

Unfolding is a common tool to correct for both mentioned detector effects. A detector simulation provides the necessary knowledge about the connection between the measured and the original values of a given observable. In the following, the original values are referred to as particle-level or generated (*Gen*) observables. They result from a simulation of the signal process as described in Section 5.3.2. Propagating the particle-level quantities through the detector simulation yields values as measured in the experiment, in the following referred to as detector-level or reconstructed (*Reco*) quantities.

Mathematically, the connection between particle-level and detector-level can be expressed by a convolution integral. In a discretized phase space, which is the usual case when binning histograms for a cross section measurement, this integral turns into a matrix multiplication. The response matrix (K_{ij}) gives the probability for a value generated in bin j of the particle-level histogram h^Gen to be reconstructed in bin i of the detector-

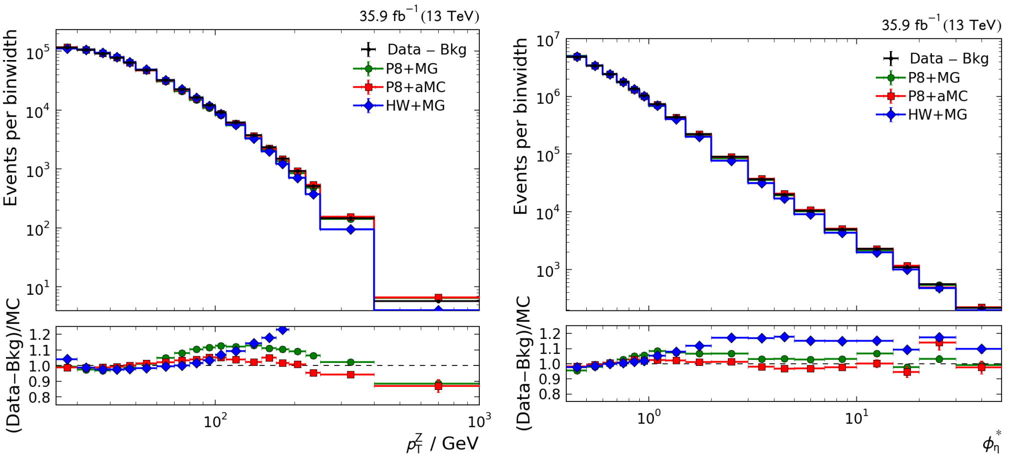


Figure 5.11: The $p_{\mathrm{T}}^{\mathrm{Z}}$ (left) and the ϕ_η^* (right) distributions on detector level derived from different simulated samples compared to the data after background subtraction. The simulated distributions are normalized to the data for a shape comparison. A sizeable difference between the HW+MG simulation and the data occurs at high $p_{\mathrm{T}}^{\mathrm{Z}}$.

level histogram h^{Reco}. Inefficiencies are included in the normalization of the response matrix K. Misidentified events, in the following referred to as *fakes*, are respected by adding a histogram h^{Fake}.

The equation to be solved can be formulated as:

$$h^{\mathrm{Reco}} = K\ h^{\mathrm{Gen}}\ +\ h^{\mathrm{Fake}} \tag{5.8}$$

There are several possibilities to solve this equation, depending on the properties of the response matrix. The easiest way to access h^{Gen} is via a simple matrix inversion, which works well if the response matrix K is sufficiently diagonal. In this case, migrations among the bins are small and appear mainly between neighbouring bins.

The condition number is an indicator of the invertibility of the response matrix. It is defined as the ratio between the largest and the smallest eigenvalue. If the condition number becomes large, a numerical matrix inversion becomes unstable and likely to create unphysical fluctuations.

This can be reduced by regularization techniques. The D'Agostini algorithm provides an iterative method[66], whereas the TUnfold framework [67] introduces an additive regularization term proportional to a parameter τ. However, regularization is not always needed. At the suggestion of the CMS statistics committee, matrix inversion without regularization is recommended to be used if the condition number of the response matrix is smaller than 10 [68].

Several simulations (DY + jets) that describe the inclusive Z (→ $\mu\mu$) + jet production have been introduced in Section 5.3.2. The distributions shown in Figures 5.6, 5.7 and 5.8 indicate that the chosen backgrounds in conjunction with the P8+MG simulation at LO precision describe the data sufficiently well. Therefore, it is used for the construction of the response matrices.

The P8+MG simulation has been favoured over the P8+aMC simulation at NLO precision because the latter suffers from the occurrence of negative event weights, leading to technical challenges in phase space regions with low statistical occupation. However, it is used as a cross check, as well as the HW+MG simulation created by HERWIG++.

A shape comparison of the simulated $p_{\mathrm{T}}^{\mathrm{Z}}$ and ϕ_{η}^{*} distributions on detector level to the data distributions is shown in Figure 5.11. The HW+MG simulation shows sizable deviations from the data. A potential bias of unfolding results which arise from these deviations is included in the systematic unfolding uncertainty introduced in Section 5.5.6.

5.5.1 Response matrices

The unfolding procedure is able to correct a measurement for inefficiency, misidentification and smearing effects by the use of a response matrix containing all the necessary information. It also makes it possible to take correlations among bins into account.

The response matrix is constructed by filling events from the simulation into a two-dimensional histogram. Only events that pass the selection both on particle-level and on detector-level are filled according to the generated bin on the x-axis and the reconstructed bin on the y-axis. Each column is normalized to the total number of generated events in the corresponding bin. Inefficiency and misidentification effects are taken into account by filling additional one-dimensional histograms with events that do not pass the particle-level or detector-level selection, respectively.

Response matrices for migrations among bins of $p_{\mathrm{T}}^{\mathrm{Z}}$, ϕ_{η}^{*}, y^{Z}, y^{jet1}, y^{*} and y_{b}, derived from the P8+MG simulation, are shown in Figure 5.12. Migrations occur mainly between neighbouring bins of ϕ_{η}^{*}, $p_{\mathrm{T}}^{\mathrm{Z}}$ and y^{Z} with a probability at percent level, migrations between further bins occur with a probability of less than one permille.

However, migrations between bins of y^{jet1} can be observed with a significant probability, even to non-neighbouring bins. These migrations propagate to y^{*} and y_{b}. They will occur if the leading reconstructed jet does not originate from the leading generated jet, but from a subleading generated jet. The order of the jets is therefore *switched* due to the detector resolution in $p_{\mathrm{T}}^{\mathrm{jet}}$. The use of additional selection criteria to reduce additional jet activity, for example kinematic cuts on the ratios $p_{\mathrm{T}}^{\mathrm{jet2}}/p_{\mathrm{T}}^{\mathrm{Z}}$ or $p_{\mathrm{T}}^{\mathrm{jet2}}/p_{\mathrm{T}}^{\mathrm{jet1}}$ leads to a significant event loss without satisfying improvement of the respective response matrix. The effect, in the following referred to as *jet switching*, is therefore treated as a detector effect, covered by unfolding and further discussed in Section 5.5.4.

The detector resolution of each observable can vary in different phase space regions. To correctly account for migrations within the triple-differential phase space, a mapping to a one-dimensional *global* binning has to be introduced. The $p_{\mathrm{T}}^{\mathrm{Z}}$ and ϕ_{η}^{*} histograms of each rapidity bin are arranged consecutively to create a combined histogram with 264 and 231 global bins, respectively. The map in Figure 5.13 illustrates the order of the rapidity bin arrangement. The rapidity bin with index 1 will be referred to as the central rapidity bin in the following sections, with index 5 as the high-y_{b} and with index 15 as the high-y^{*} bin. Resulting response matrices for this approach, derived from the P8+MG

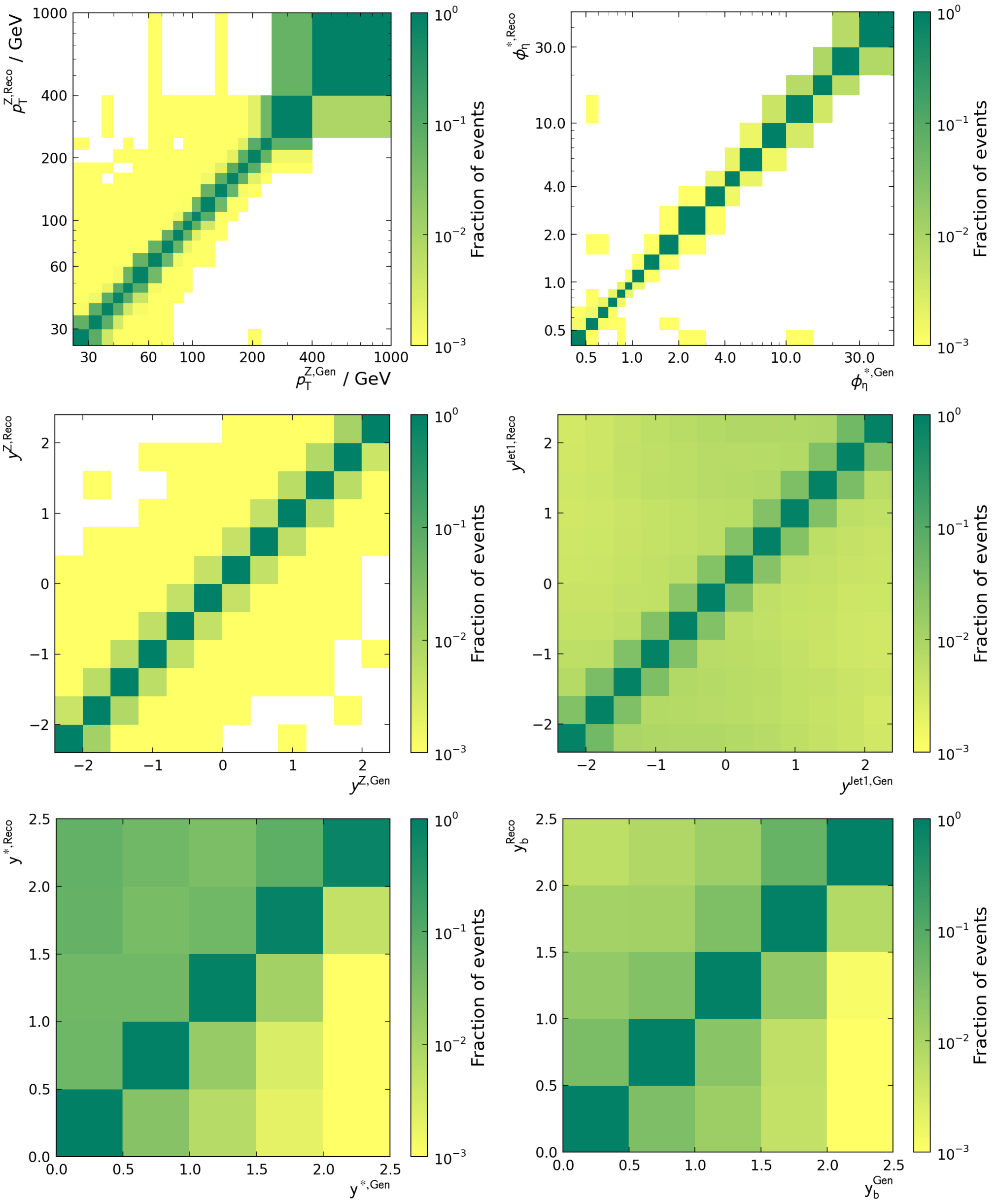


Figure 5.12: Response matrices for migrations among bins of p_T^Z (top left), ϕ_η^* (top right), y^Z (centre left), y^{jet1} (centre right), y^* (bottom left) and y_b (bottom right) derived from the P8+MG simulation.

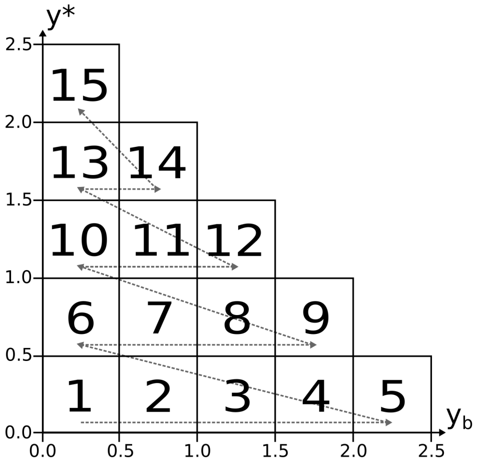


Figure 5.13: Order of the rapidity bin arrangement when mapping the triple differential phase space to a one-dimensional histogram. The rapidity bin with index 1 is called the central rapidity bin, with index 5 the high-y_b and with index 15 the high-y^* bin.

simulation, are shown in Figure 5.14.

Response matrices derived from full detector simulations can suffer from a limited number of events. For this reason, a so-called *forward smearing* approach is used for this analysis. Inefficiency, misidentification and resolution parameters are estimated from the simulation and the response matrices are filled using a toy Monte Carlo simulation. The advantage resulting from this approach is that the statistical uncertainty due to insufficient bin occupations can be reduced to a negligible level. On the other hand, systematic uncertainties are introduced to reflect the dependence on the estimation methods, which are explained in the following sections.

5.5.2 Resolution

To estimate the resolutions, the generated and the reconstructed values of a given observable are compared separately in each bin on an event-by-event basis. Their difference, normalized to the generated value, is filled into a histogram to derive the relative resolution of p_T^Z and ϕ_η^*.

$$R(p_\mathrm{T}^\mathrm{Z}) = \frac{p_\mathrm{T}^\mathrm{Z,Reco} - p_\mathrm{T}^\mathrm{Z,Gen}}{p_\mathrm{T}^\mathrm{Z,Gen}} \tag{5.9}$$

$$R(\phi_\eta^*) = \frac{\phi_\eta^{*,\mathrm{Reco}} - \phi_\eta^{*,\mathrm{Gen}}}{\phi_\eta^{*,\mathrm{Gen}}} \tag{5.10}$$

Since y^* and y_b are defined as combinations of rapidities of objects with different resolution characteristics, the initial observables y^Z and y^jet1 are used. It is not reasonable to

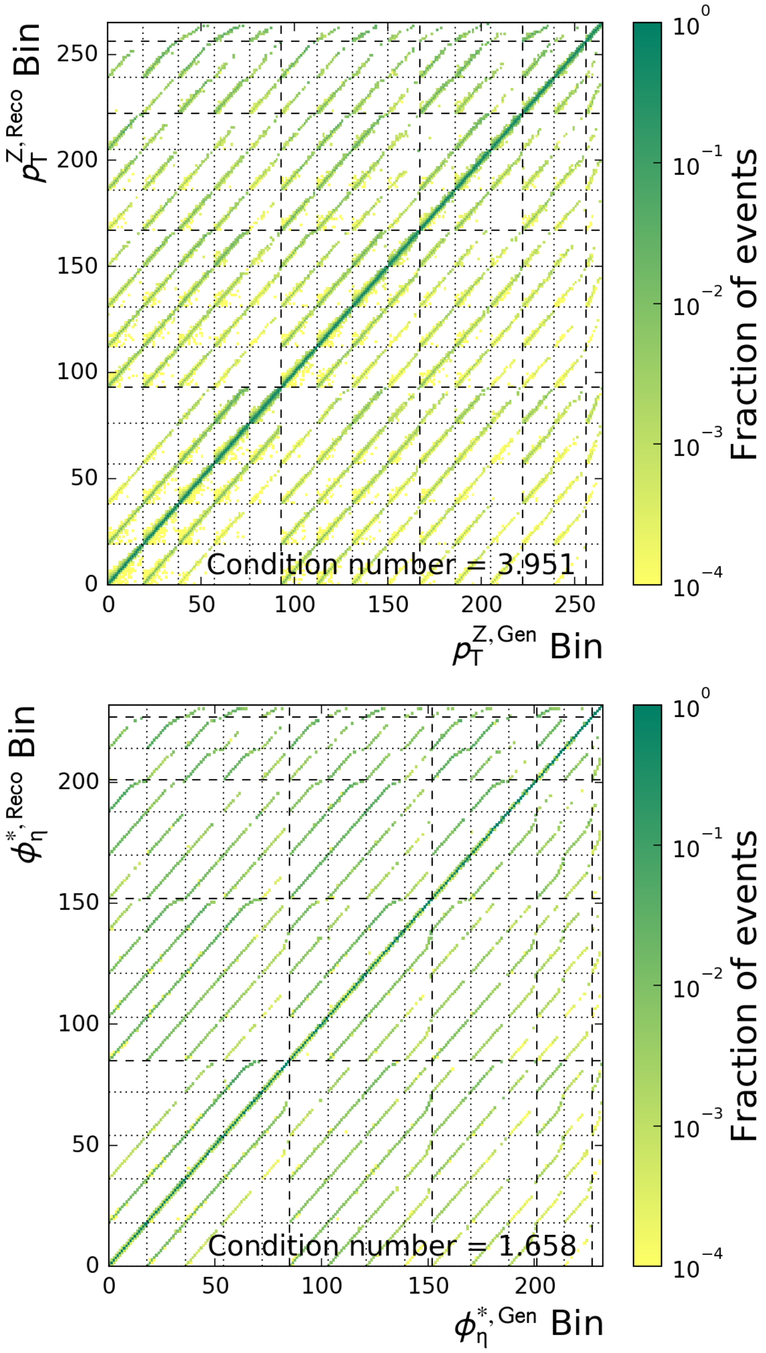


Figure 5.14: Response matrices for triple differential unfolding in $(p_{\mathrm{T}}^{\mathrm{Z}}, y^{*}, y_{\mathrm{b}})$ (top) and $(\phi_{\eta}^{*}, y^{*}, y_{\mathrm{b}})$ (bottom) derived from the P8+MG simulation.

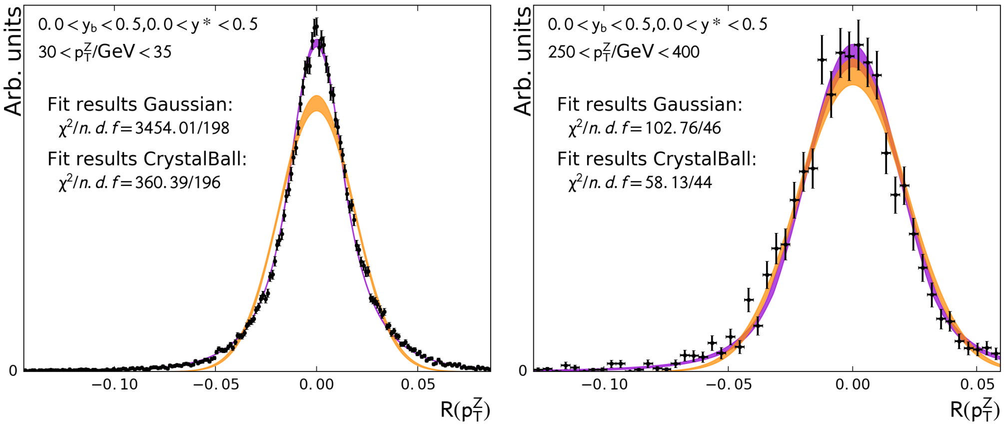


Figure 5.15: Examples of resolution histograms of p_T^Z at low (left) and at high transverse momentum (right) in the central rapidity bin. A Gaussian fit (orange) can not fully describe the shapes, especially at low energies. A symmetrized Crystal Ball function (violet) is able to describe the non-Gaussian tails.

calculate a relative resolution of a quantity distributed over a range which contains zero, therefore absolute instead of relative resolutions are derived.

$$R(y^{\mathrm{Z}}) = y^{\mathrm{Z,Reco}} - y^{\mathrm{Z,Gen}} \tag{5.11}$$

$$R(y^{\mathrm{jet1}}) = y^{\mathrm{jet1,Reco}} - y^{\mathrm{jet1,Gen}} \tag{5.12}$$

To avoid bias in the resolution of y^{jet1} originating from jet switching, reconstructed jets are matched to generated jets within an angular distance of $\Delta R < 0.2$. The rapidity of the matched generated jet is taken into account for the calculation of $R(y^{\mathrm{jet1}})$.

Histograms of $R(p_{\mathrm{T}}^{\mathrm{Z}})$ or $R(\phi_\eta^*)$, $R(y^{\mathrm{Z}})$ and $R(y^{\mathrm{jet1}})$ are filled for each bin of $(p_{\mathrm{T}}^{\mathrm{Z}}, y^*, y_{\mathrm{b}})$ or $(\phi_\eta^*, y^*, y_{\mathrm{b}})$. The distributions are truncated to 98.5 % of the total histogram content to avoid bias due to outliers. Figure 5.15 shows examples of the resulting histograms.

Three kinds of resolution measures are derived from each histogram: the root mean square (RMS), the Gaussian width and the Crystal Ball width.

The RMS as a measure of the width of the distribution serves as an upper limit to the resolution.

The fit of a Gaussian function can not describe the shape of the distribution, although its width yields an approximation to the resolution.

The fit of a symmetrized Crystal-Ball function is able to describe the non-Gaussian tails and its width yields a lower limit to the resolution. However, in contrast to the relatively robust fit convergence of the Gaussian function, the fit convergence of the Crystal Ball function can not be guaranteed without biasing the parameters by limiting their range.

Due to its straightforward evaluation and its characteristic as an upper bound to the resolution, the RMS has been chosen as resolution parameter, attached with an RMS

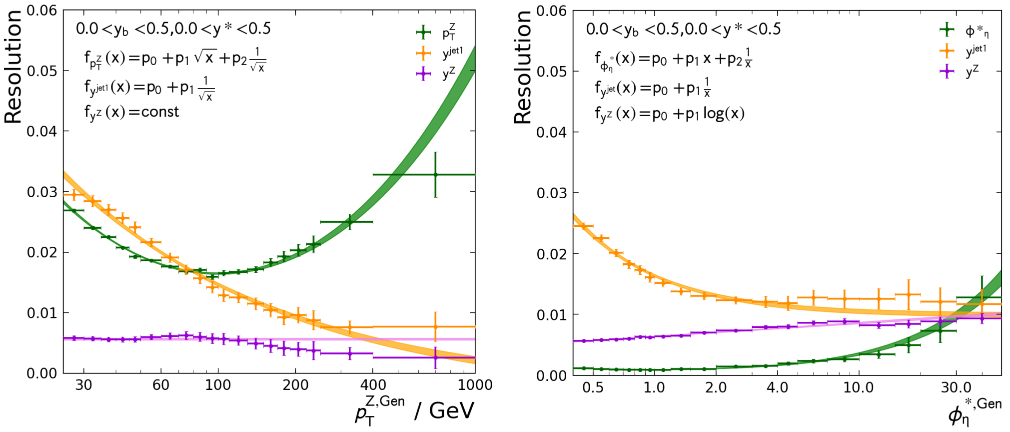


Figure 5.16: The RMS results of histograms filled with $R(p_\mathrm{T}^\mathrm{Z})$ or $R(\phi^*_\eta)$, $R(y^\mathrm{Z})$ and $R(y^\mathrm{jet1})$ in bins of p_T^Z (left) and ϕ^*_η (right) in the central rapidity bin. Different parametrizations have been chosen to smoothen the p_T^Z and ϕ^*_η dependence. The RMS uncertainties are scaled to account for bias by the chosen parametrization and non-Gaussian behaviour of the distributions. An overview of all rapidity bins can be found in Figures A.3 and A.4.

uncertainty. Figure 5.16 displays its results for p_T^Z or ϕ^*_η, y^Z and y^jet1 as a function of p_T^Z and ϕ^*_η in the central rapidity bin.

The dependence on p_T^Z or ϕ^*_η is parametrized by suitable functions to reduce unphysical fluctuations. The functions have been chosen separately for each resolution to emulate the p_T^Z and ϕ^*_η dependence.

The relative resolution of p_T^Z or ϕ^*_η exhibits a minimum that is emulated by the additive combination of two power law functions with fixed opposite-sign exponents. The observed trend can be understood as the combination of two competing effects that limit the muon resolution: their momentum resolution worsens at high momenta because of the reduced track curvature that is taken into account for its estimation; their angular resolution worsens at low momenta because the track extrapolation to the interaction vertex becomes less accurate. The resolution of p_T^Z or ϕ^*_η, observables derived as a combination of muon components, is therefore determined by the limiting component.

The resolution of y^jet1 increases with the transverse momentum $p_\mathrm{T}^\mathrm{jet1}$, which is correlated to p_T^Z and ϕ^*_η. A power law with a fixed negative exponent is used for the parametrization.

The resolution of y^Z follows an approximately logarithmic increase with ϕ^*_η. Its dependence on p_T^Z can not be parametrized by a simple function. However, its values are small compared to the resolution of p_T^Z and y^jet1, so that a constant function suffices for its parametrization.

Since the distributions used for the RMS calculation are not Gaussian, its uncertainties are only approximate values. An underestimation of uncertainties potentially results in large results of the goodness-of-fit estimator ($\chi^2/n.d.f.$). To account for both underestimation of uncertainties and bias by the choice of the parametrization function, the

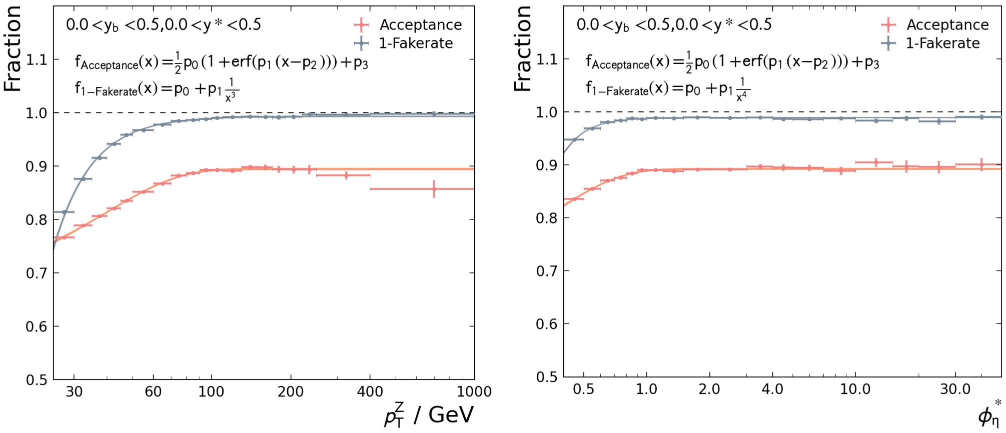


Figure 5.17: Acceptance and fakerate in bins of $p_{\mathrm{T}}^{\mathrm{Z}}$ (left) and ϕ_η^* (right) in the central rapidity bin. Parametrizations have been chosen to smoothen the $p_{\mathrm{T}}^{\mathrm{Z}}$ and ϕ_η^* dependence. The binomial uncertainties are scaled to account for bias by the chosen parametrization and deviations from the binomial approximation. An overview of all rapidity bins can be found in Figures A.5 and A.6.

uncertainties are rescaled uniformly such that the goodness-of-fit estimator becomes equal to unity. The resulting parametrization and rescaled uncertainties are shown in Figure 5.16.

5.5.3 Efficiencies

In contrast to the response matrix filled with events that pass the selection both on particle level and on detector level, there are events that pass only one of them. Unfolding is able to correct for this sort of detector effects by consideration of marginal distributions. In this context, an event is categorized as *loss* if it passes the particle-level selection, but not the detector-level selection. Conversely, an event is categorized as *fake* if it passes the detector-level selection, but not the particle-level selection.

The derived parameters are called acceptance A and fakerate F and are estimated in every phase space bin. The acceptance A is calculated as the fraction of events generated in the respective bin that are reconstructed in *any* bin.

$$A = \frac{\#\{\text{event in } \mathbf{Gen} \text{ bin } \& \text{ event in } \textit{any} \ \mathbf{Reco} \text{ bin}\}}{\#\{\text{event in } \mathbf{Gen} \text{ bin}\}} \tag{5.13}$$

The fakerate F is calculated as the complementary fraction of events reconstructed in the respective bin that have been generated in *any* bin.

$$F = 1 - \frac{\#\{\text{event in } \mathbf{Reco} \text{ bin } \& \text{ event in } \textit{any} \ \mathbf{Gen} \text{ bin}\}}{\#\{\text{event in } \mathbf{Reco} \text{ bin}\}} \tag{5.14}$$

It is worth noting that the fakerate is determined as a function of the detector-level

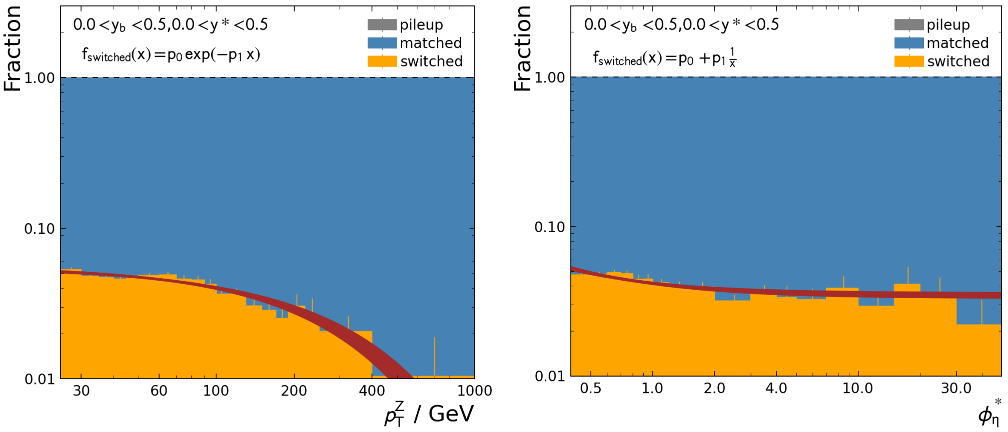


Figure 5.18: Fraction of matched, switched and pileup events in bins of p_T^Z (left) and ϕ^*_η (right) in the central rapidity bin. The switching fraction is parametrized by a suitable function. Its binomial uncertainties are scaled to account for bias by the chosen parametrization and deviations from the binomial approximation. An overview of all rapidity bins can be found in Figures A.7 and A.8.

values of p_T^Z or ϕ^*_η, y^* and y_b, the acceptance as a function of their particle-level values. Figure 5.17 shows the results for the central rapidity bin.

The results have been parametrized with suitable functions. Whereas an error function proved to be a good choice for the acceptance whose shape resembles a turn-on curve, a hyperbolic function appears to be the better choice to describe the fakerate.

Uncertainties are estimated as binomial errors, which is a good approximation if the rates are not too close to zero or unity. As the fakerate F approaches zero at higher p_T^Z or ϕ^*_η (corresponding to values of $1 - F$ in proximity to unity), the uncertainties are underestimated and uniformly scaled such that the goodness-of-fit estimator ($\chi^2/n.d.f.$) becomes equal to unity. The resulting parametrization and rescaled uncertainties are added to Figure 5.17.

5.5.4 Switching

The study of the jet rapidity resolution revealed an additional challenge following from the fact that jets are indistinguishable objects. The analysis takes into account only one jet, although the occurrence of additional jets is very likely. To avoid ambiguity, the relevant jet is chosen to be the one with the largest transverse momentum.

However, the jet ordering can differ if one compares an event topology on detector level to the one on particle level within a simulation. This effect was found to be small in inclusive jet and dijet cross section measurements, bypassed by the use of angular jet matching in the former and the averaged dijet transverse momentum in the latter case.

The substitution of the leading particle-level jet by a subleading particle-level jet that is reconstructed as the leading detector-level jet immediately propagates to the determination of the generated and the reconstructed values of y^* and y_b in an inclusive Z + jet cross section measurement. For a quantitive characterization of the effect, reconstructed events passing the selection are tagged as one of three exclusive categories that are based on jet matching:

- **matched**: An event in which the generated leading jet and the reconstructed leading jet have an angular distance of $\Delta R < 0.2$ is tagged as *matched*.
- **switched**: An event in which there is a *sub*leading generated jet with an angular distance of $\Delta R < 0.2$ to the reconstructed jet is tagged as *switched*.
- **pileup**: An event in which there is *no* generated jet with an angular distance of $\Delta R < 0.2$ to the reconstructed jet is tagged as *pileup*. Note that an event tagged as pileup does not necessarily imply the reconstructed leading jet to be an actual pileup jet, it can as well originate from a generated jet beyond the accessible phase space.

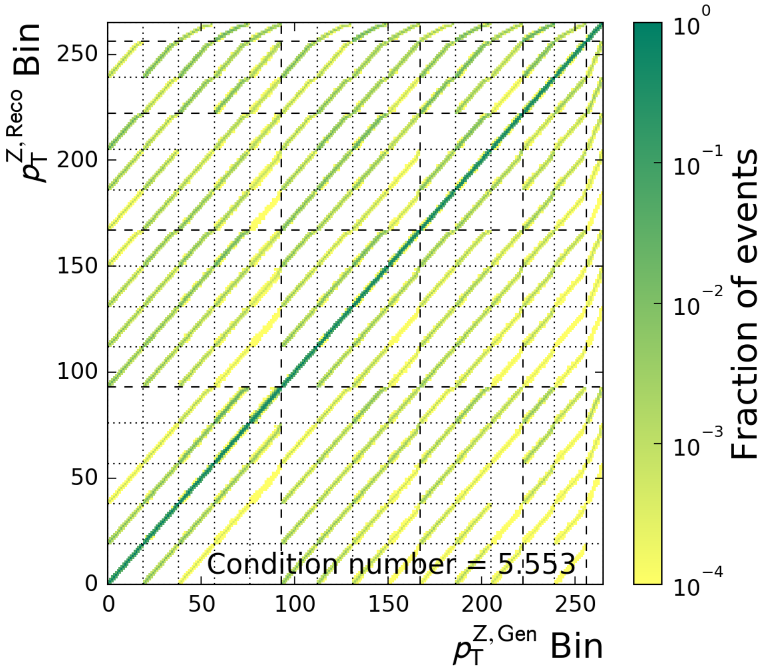
Figure 5.18 shows the fractions of events that fall in the respective categories in the central rapidity bin. While pileup tends to be a negligible effect, the fraction of switched events lies in the range of less than 5% in the central rapidity bin, increasing to up to 20% in higher rapidity regions. The effect is corrected for by unfolding, considering switching to be a migration to a distant bin. The switching fractions are parametrized by suitable functions, namely an exponential function for the p_T^Z and a hyperbolic function for the ϕ_η^* dependence. To account for the underestimation of binomial uncertainties and bias by the parametrization, the uncertainties attached to each value of the switching fraction are again uniformly scaled such that the goodness-of-fit estimator ($\chi^2/n.d.f.$) becomes equal to unity. The resulting parametrization and rescaled uncertainties are added to Figure 5.18.

5.5.5 Forward smearing

Based on the parameters derived in the previous sections, response matrices for the triple differential unfolding in ($p_\mathrm{T}^\mathrm{Z}, y^*, y_\mathrm{b}$) and ($\phi_\eta^*, y^*, y_\mathrm{b}$), respectively, are filled using a toy Monte Carlo simulation. This approach, also called forward smearing, accounts for fluctuations of the response matrix in phase space regions with low statistical occupation.

The dependence of the parameters on p_T^Z or ϕ_η^*, estimated separately in each rapidity bin, is mapped to the one-dimensional global binning as illustrated in Figure 5.13. The forward smearing procedure can then be split into four steps:

1. A toy event is generated by random values of p_T^Z or ϕ_η^*, y^Z and y^jet1, following the particle-level distribution of the full simulation. (p_T^Z or $\phi_\eta^*, y^*, y_\mathrm{b}$) is calculated and mapped to the generated global bin i.
2. Switching occurs with a probability S, which is given by the switching fraction in bin i. A random number determines if the generated value of y^jet1 is replaced according to a distribution which is derived from the full simulation and respects



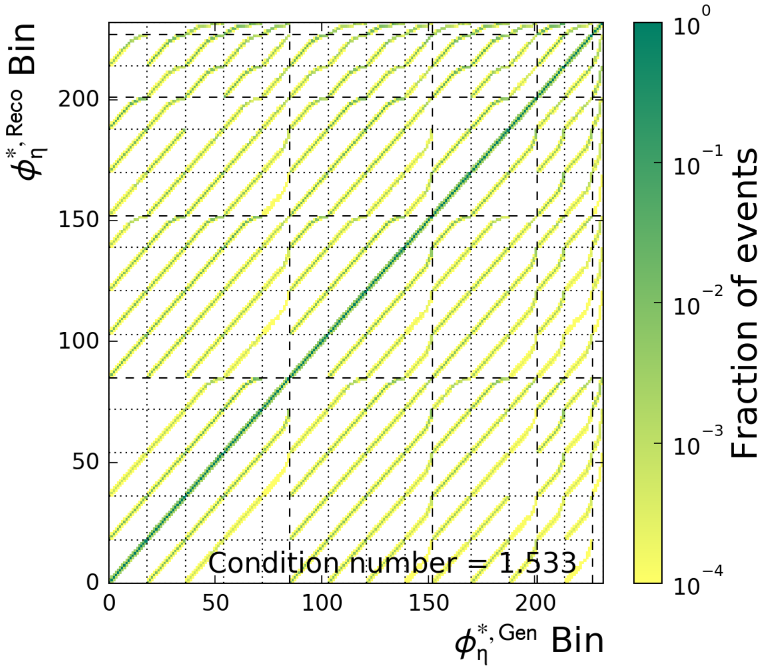


Figure 5.19: Response matrices for triple differential unfolding of (p_T^Z, y^*, y_b) (top) and (ϕ_η^*, y^*, y_b) (bottom), created by forward smearing.

its correlation to y^{Z}. The resulting ($p_{\mathrm{T}}^{\mathrm{Z}}$ or $\phi_{\eta}^{*}, y^{*}, y_{\mathrm{b}}$) is mapped to an intermediate global bin k.

3. Smearing is performed by manipulating each value of $p_{\mathrm{T}}^{\mathrm{Z}}$ or ϕ_{η}^{*}, y^{Z} and y^{jet1} independently by a smearing parameter s. $p_{\mathrm{T}}^{\mathrm{Z}}$ or ϕ_{η}^{*} are multiplied by $s = 1 + Rr$, where R is the relative resolution of the observable in bin k and r a normally distributed random number. For y^{Z} and y^{jet1}, $s = Rr$ is added to the respective value, where R is the absolute resolution of the observable in bin k and r a normally distributed random number. The resulting values of ($p_{\mathrm{T}}^{\mathrm{Z}}$ or $\phi_{\eta}^{*}, y^{*}, y_{\mathrm{b}}$) are mapped to the reconstructed global bin j.
4. Acceptance and fakerate are simulated independently. The event is filled into the reconstructed marginal histogram bin j only with a probability A, given by the acceptance in bin k. Likewise, it is filled to the generated marginal histogram bin i with a probability of $1 - F$, where F is the fakerate in bin j. If both histograms are filled, the event is filled into the response matrix bin (i,j) as well.

The response matrices based on 10^9 toy events, which is a factor of 100 in comparison to the number of simulated events, are shown in Figure 5.19. Disregarding sparsely populated off-diagonal areas, they are in good agreement with the response matrices derived from the full simulation in Figure 5.14, but with significantly reduced statistical fluctuations.

Quantitative benchmarks evaluating how well the method works and a study of uncertainties, both associated with statistical fluctuations and systematic effects due to the parameter estimation, is described in the following section.

5.5.6 Closure test

The response matrices, both those derived from the simulation (Figure 5.14) and those from forward smearing (Figure 5.19) have been found to have condition numbers well below 10. The TUnfold framework without regularization can be used to perform the unfolding and has been checked to yield the same results as the matrix inversion method. Background contributions are subtracted from the data before unfolding to obtain signal distributions that can be compared to the corresponding simulation.

Distributions derived from the full simulations are used to obtain a benchmark for the closure of an unfolding method. The reconstructed distributions contained in the simulated samples serve as pseudodata and are used as an input to the unfolding procedure. The unfolded output is compared to the generated distributions of the respective simulation. Their difference is then compared to the uncertainties associated with the respective method to obtain a quantitive measure for the accuracy of the method. An unfolding method can inherit statistical and systematic uncertainties.

Statistical unfolding uncertainties

To estimate a potential bias due to the limited bin occupation of a response matrix, a set of 1000 pseudoexperiments is performed. For every pseudoexperiment, each bin of the response matrix (including fake and loss distributions) is shifted by its statistical uncertainty multiplied by a normally distributed random number.

The data distributions are used to estimate the effect of this variation. After subtracting the backgrounds, they are unfolded with the altered response matrix and the result is compared to the central result unfolded with the unaltered response matrix.

The root mean square of the deviations of each pseudo experiment from the central result is taken as the statistical unfolding uncertainty.

Systematic unfolding uncertainties

The construction of response matrices by forward smearing depends on the estimation of the characteristic parameters resolution, acceptance, fakerate and switching fraction. Their dependence on $p_{\mathrm{T}}^{\mathrm{Z}}$ and ϕ_{η}^{*} has been parametrized to smoothen unphysical fluctuations. The 68% (1σ) confidence interval of the resulting parametrization, evaluated at each bin centre, defines the smoothened parameter uncertainties.

These uncertainties are propagated through the unfolding procedure by creating response matrices with the respective parameter simultaneously shifted upwards and downwards by its uncertainty in each bin. After subtracting the backgrounds, the data distributions are unfolded with each of the resulting response matrices and compared to the central result. The maximum difference between the respective upwards/downwards variations and the central result is taken as the uncertainty on the resulting distribution associated with the parameter in question.

In principle, the parameters can be derived from any of the simulation samples described in Section 5.3.2 and compared in Figure 5.11. In practice, the P8+aMC simulation proved to be technically challenging due to the occurrence of negative event weights and has not been used for a toy response creation. The P8+MG simulation with the highest number of simulated events serves as the reference and has been used to illustrate the previous results. The same procedure has been used for the HW+MG simulation containing about half as many events within the relevant phase space.

Although the same detector simulation was used in the simulated samples, the implementation of parton shower and hadronization models in the event generators PYTHIA8 and HERWIG++ can lead to differences of the parameters beyond statistical fluctuations. Therefore, the difference between the unfolding result of the background-subtracted data distributions using a toy response matrix based on the P8+MG sample and the unfolding result of the same histograms using a toy response matrix based on the HW+MG sample is taken as an additional uncertainty source.

The uncertainty sources are assumed to be uncorrelated, such that their quadratic sum is taken as the systematic unfolding uncertainty.

χ^2-test

The closure of an unfolding method is characterized by a comparison of an unfolded reconstructed distribution to the respective generated distribution derived from the same

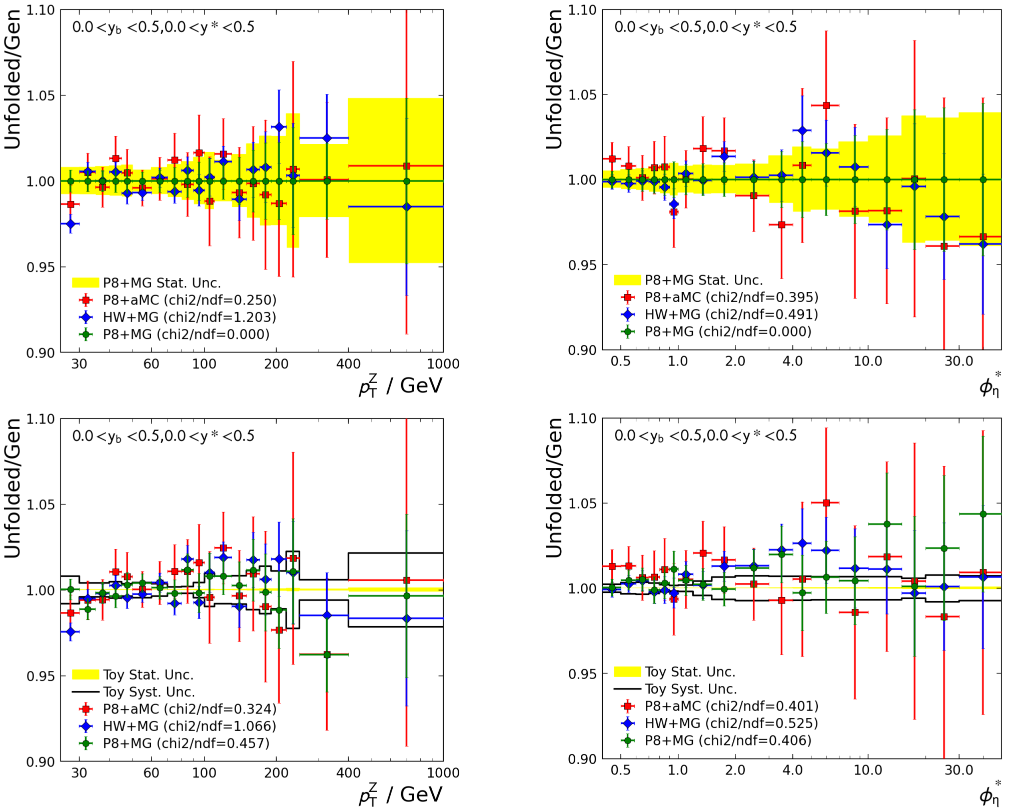


Figure 5.20: Ratio of the unfolded reconstructed to the respective generated p_T^Z (left) and ϕ_η^* (right) distribution derived from the same simulation in the central rapidity bin. The unfolding has been carried out with response matrices filled from the P8+MG simulation (top) or filled by forward smearing (Toy) based on the P8+MG simulation (bottom). The yellow band illustrates the statistical unfolding uncertainty, the black lines illustrate the systematic unfolding uncertainty (where applicable) and the error bars illustrate the statistical uncertainty of the unfolded distributions.

simulation. Their ratio for the three different simulation inputs is shown in Figure 5.20. The central rapidity bin is chosen for the illustration.

The response matrix used in the top row of Figure 5.20 has been filled from the P8+MG simulation, the response matrix used in the bottom row via forward smearing based on the P8+MG simulation. A comparison of the two figures illustrates the improved precision when using the forward smearing approach. Statistical unfolding uncertainties, following from a limited response matrix occupation, are shown as a yellow band which is reduced to a negligible contribution. The systematic uncertainties, following from the parameter estimation, shown as a black line in the bottom row of Figure 5.20, have been found to be smaller than the statistical uncertainty in the top row of Figure 5.20.

To ensure the applicability of the method, a $\chi^2/n.d.f.$ value was calculated separately in each rapidity bin and for each simulated sample to obtain a quantitative characteristic

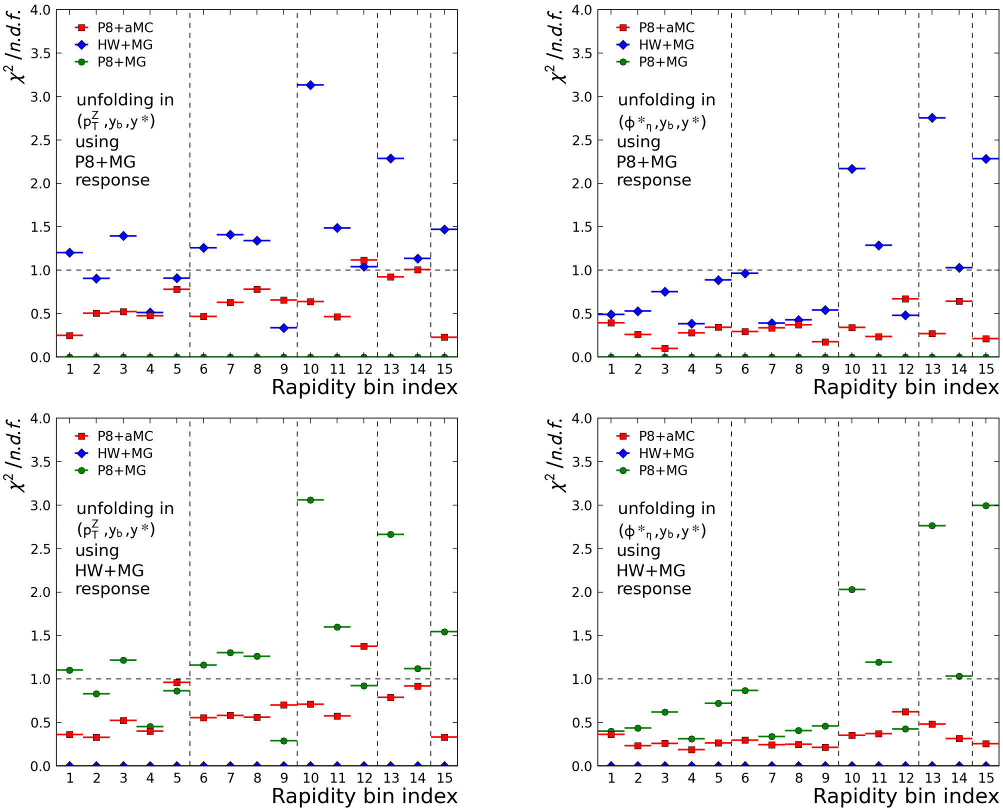


Figure 5.21: Values of $\chi^2/n.d.f.$ determined for each rapidity bin, for unfolding in (p_T^Z, y^*, y_b) (left) and (ϕ^*_η, y^*, y_b) (right). The unfolding has been carried out with a response matrix filled from the P8+MG simulation (top) or from the HW+MG simulation (bottom).

for the closure. χ^2 is calculated as the squared difference of the unfolded reconstructed histogram h^{Unf} to the generated histogram h^{Gen}, divided by the squared sum of the statistical unfolding uncertainty $\sigma_{\text{Stat.Unf.}}$, the systematic unfolding uncertainty $\sigma_{\text{Syst.Unf.}}$ and the statistical uncertainty $\sigma_{\text{Stat.MC}}$ of the unfolded distribution:

$$\chi^2 = \sum_{\text{bins}} \frac{\left(h^{\text{Unf}} - h^{\text{Gen}}\right)^2}{\sigma^2_{\text{Stat.Unf.}} + \sigma^2_{\text{Syst.Unf.}} + \sigma^2_{\text{Stat.MC}}} \tag{5.15}$$

Note that $\sigma_{\text{Syst.Unf.}} = 0$ if the response matrix is filled from the simulations.

χ^2 is divided by $n.d.f.$, given as the number of p_T^Z or ϕ^*_η bins within the respective rapidity bin. The closure is accepted to be *good* if the resulting value of $\chi^2/n.d.f.$ is approximately equal to unity. The values as a function of the rapidity bin index (see Figure 5.13) are presented in Figure 5.21 and 5.22.

For Figure 5.21, the response matrix have been filled from the simulation, either P8+MG or HW+MG. As expected, unfolding of a pseudodata distribution using a response

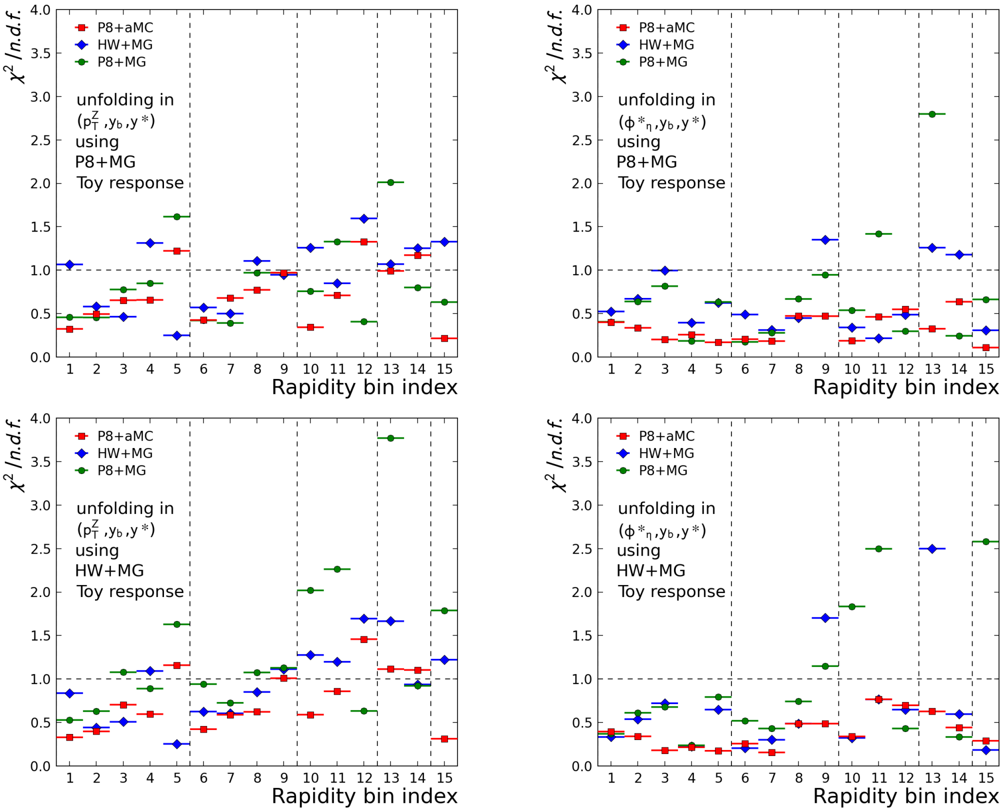


Figure 5.22: Values of $\chi^2/n.d.f.$ determined for each rapidity bin, for unfolding in $(p_\mathrm{T}^\mathrm{Z}, y^*, y_\mathrm{b})$ (left) and $(\phi^*_\eta, y^*, y_\mathrm{b})$ (right). The unfolding has been carried out with a response matrix filled by forward smearing based on the P8+MG simulation (top) or on the HW+MG simulation (bottom).

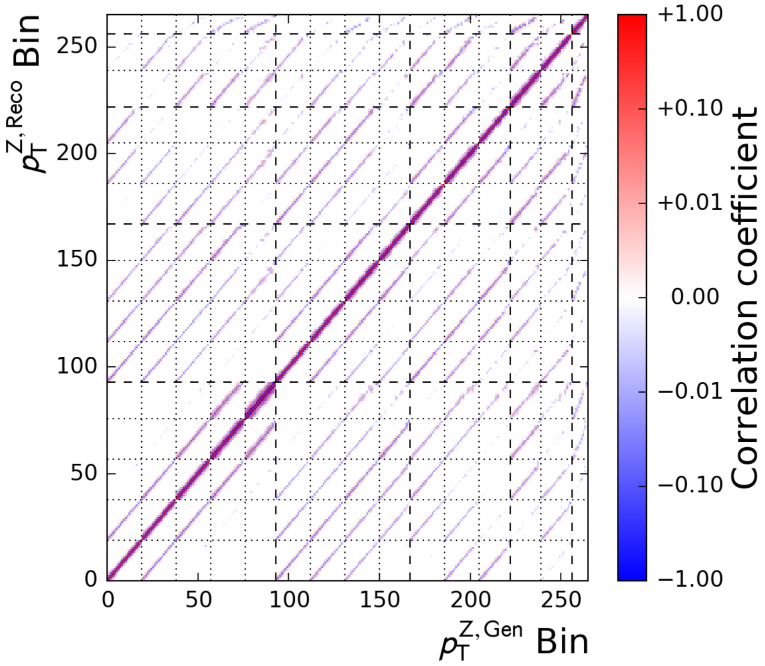
matrix from the same simulation yields exactly the generated distribution and therefore $\chi^2\ /n.d.f. = 0$.

The P8+aMC pseudodata inherits large statistical uncertainties (see also Figure 5.20) as a consequence of the occurrence of negative event weights. Therefore, the resulting values of $\chi^2/n.d.f.$ tend to be small ($\chi^2/n.d.f. \lesssim 1$) for any unfolding method.

If the P8+MG pseudodata is unfolded with a response matrix from the HW+MG simulation (or vice versa), unusually large values of $\chi^2/n.d.f.$ (2.0 to 3.5) are reached consistently in certain rapidity bins. This can be explained by differences in the Monte Carlo modelling.

For Figure 5.22, the response matrices have been filled by forward smearing. The parameters were derived either from the P8+MG or the HW+MG simulation.

The figure indicates that unfolding by forward smearing based on the P8+MG simulation is compatible with all pseudodata distributions ($\chi^2/n.d.f.$ mostly between 0.5 and 1.5), whereas forward smearing based on the HW+MG simulation yields large values of



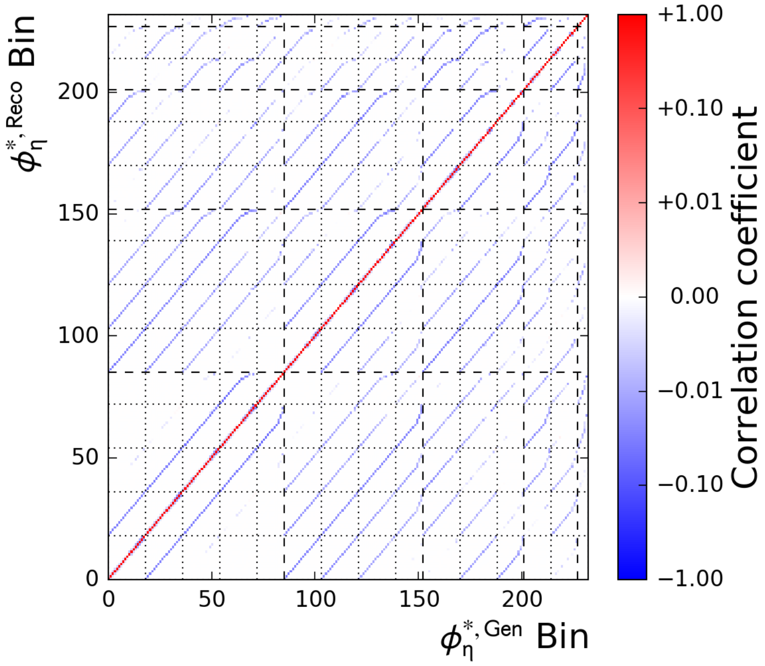


Figure 5.23: Bin-to-bin correlation matrices for triple differential unfolding of (p_T^Z, y^*, y_b) (top) and (ϕ_η^*, y^*, y_b) (bottom) using response matrices filled by forward smearing based on the P8+MG simulation. Neighbouring bins are weakly anticorrelated.

$\chi^2/n.d.f.$ in several rapidity bins.

However, even in the former case, there is one rapidity bin (#13 in ϕ^*_η unfolding) with an outlying value of $\chi^2/n.d.f. \approx 2.8$. The bin has been checked further for anomalous behaviour. As no significant effects have been observed, the large value in this bin is interpreted as a statistical fluctuation.

As a result of the closure tests, the forward smearing method based on the P8+MG simulation has been chosen as preferred unfolding method. The difference to the HW+MG simulation is taken into account by the systematic unfolding uncertainty, therein forming the dominant contribution, whereas the statistical unfolding uncertainty is negligible.

The resulting bin-to-bin correlation matrices associated with unfolding have been found to be almost diagonal. Small negative off-diagonal correlation coefficients indicate weak anticorrelations between neighbouring bins. The matrices are presented in Figure 5.23.

5.6 Experimental uncertainties

The triple differential inclusive Z ($\rightarrow \mu\mu$) + jet cross section measurements involve several experimental uncertainty sources. The following sources have been taken into account:

- **Luminosity:**

  To determine the cross sections, the measured distributions are normalized to the delivered luminosity of $35.9\,\mathrm{fb}^{-1}$ in 2016. This value was estimated from machine parameters and the collider beam. A constant relative uncertainty of 2.5% on any cross section measurement is recommended [69].

- **Background:**

  The measured distributions are extracted by subtracting the simulated background from the data distributions. This subtraction depends on the choice of the background processes and their properties. As a conservative estimate, 50% of the full background contribution in every observable bin has been used as uncertainty, leading only to about 0.5% relative uncertainty in the cross section measurement due to the small background contribution of the inclusive Z ($\rightarrow \mu\mu$) + jet process.

- **Efficiency:**

  Muon efficiency scale factors have been used on data distributions to ensure that the selection efficiencies of identification, isolation and trigger criteria agree with the simulation. The efficiencies, estimated via tag-and-probe by the CMS collaboration, were provided including systematic uncertainties on identification and isolation scale factors, dependent on the kinematic properties of the muons, and a constant uncertainty of 0.5% on the single-muon trigger efficiency [65]. The efficiencies and scale factors have been independently shifted upwards and downwards by their uncertainties and propagated through the analysis by calculating weights as explained in Section 5.4.1. The respective upwards and downwards deviations from the nominal measured distribution in each observable bin after unfolding have been

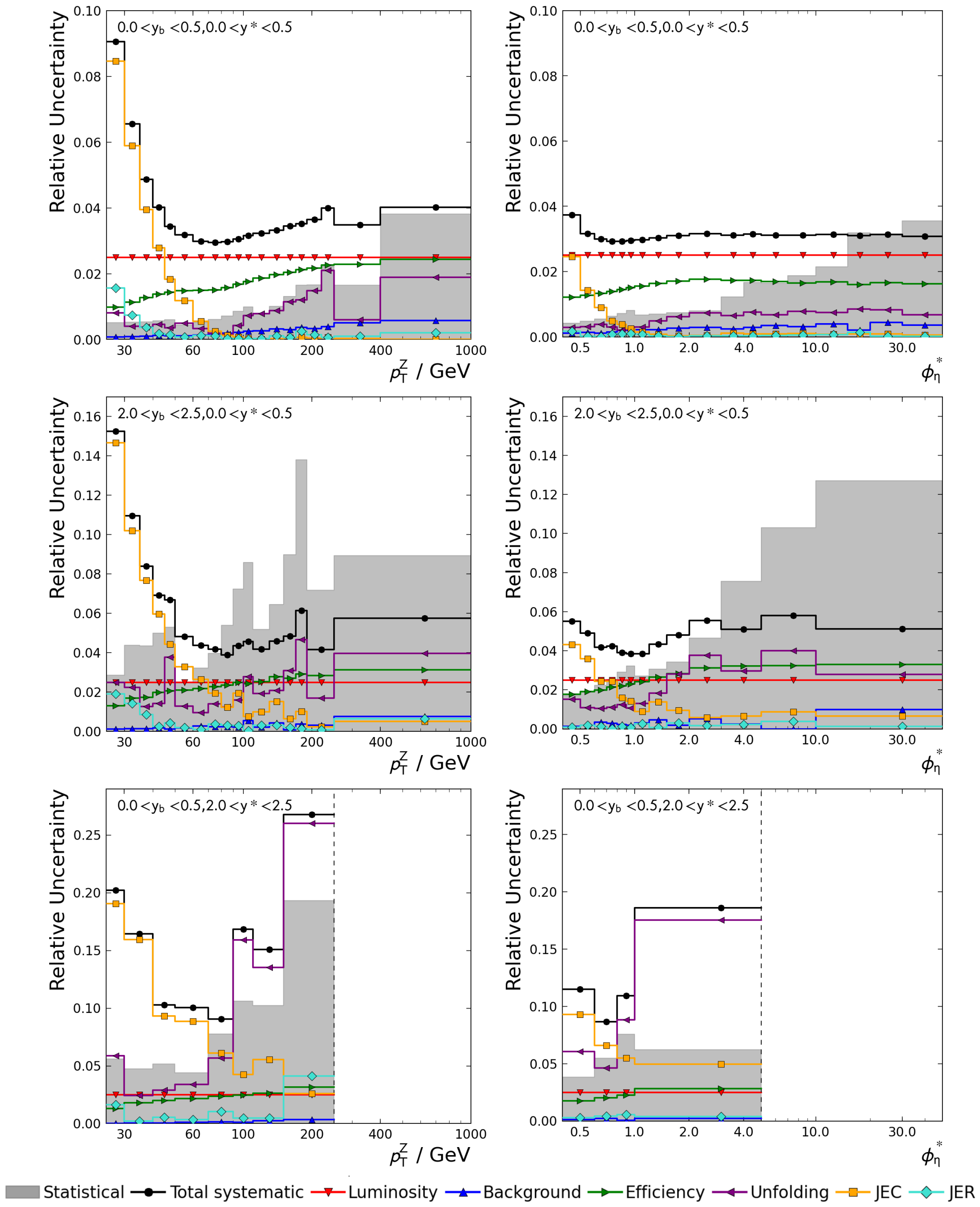


Figure 5.24: Total systematic uncertainty (black) and the different uncertainty sources (coloured) as a function of p_T^Z (left) and ϕ_η^* (right) in the central rapidity (top), the high-y_b (centre) and the high-y^* bin (bottom), together with the statistical uncertainty (light grey). The uncertainties have been symmetrized by taking half of the spread in each uncertainty source. An overview of all rapidity bins can be found in Figures A.9 and A.10.

taken as uncertainties for identification, isolation and trigger. The respective uncertainties have been assumed uncorrelated and therefore been added in quadrature to obtain the resulting efficiency uncertainty, which slightly increases with p_T^Z or ϕ_η^* and towards the high-rapidity bins.

- **Unfolding:**

  In Section 5.5.6, statistical and systematic unfolding uncertainties have been estimated. Their squared sum has been taken as the unfolding uncertainty. It is dominated by the systematic uncertainty when using the forward smearing method to fill the response matrix, which itself is dominated by the difference between the simulations that are investigated for the unfolding study. The increase with p_T^Z or ϕ_η^* and towards higher-rapidity regions, especially the high-y^* region, is closely connected to the lack of statistical occupation in these regions, complicating the parameter estimation for the forward smearing procedure.

- **JEC:**

  An uncertainty on applying jet energy corrections has been estimated by consistently varying the p_T of each jet upwards and downwards according to its jet energy calibration uncertainty, which was derived and provided by the CMS collaboration, as discussed in Chapter 4. The respective upwards and downwards deviations from the nominal measured distribution in each observable bin after unfolding have been taken as the JEC uncertainty. Since a dependence on $p_\mathrm{T}^\mathrm{jet}$ was taken into account exclusively by the selection criterion $p_\mathrm{T}^\mathrm{jet1} > 20\,\mathrm{GeV}$, only low p_T^Z or ϕ_η^* regions were found to be significantly affected by the JEC uncertainty. At low p_T^Z, it even becomes the dominant uncertainty source.

- **JER:**

  Similarly to the muon efficiencies, it was found that the jet energy resolution differs between data and simulation. To improve their agreement, the transverse momenta of jets in simulations can be additionally smeared, either using a *scaling* or a *stochastic* method. For the scaling method, a scale factor is applied on the difference between the transverse momenta of a detector-level and its matching particle-level jet, if existent. For the stochastic method, the scale factor is applied on the simulated jet energy resolution of the detector-level jet. Scale factor and resolution have been provided by the CMS collaboration, which also recommends the use of a *hybrid* method: the scaling method is applied if a matching particle-level jet can be identified, else the stochastic method is applied [70]. Toy response matrices for unfolding have been created by forward smearing, as explained in Section 5.5.5, based on a P8+MG simulation on which the hybrid smearing method was applied. The JER uncertainty is taken as the difference between the measured distributions in each observable bin after unfolding with and without additional smearing. It was found that the resolution of y^jet1 as well as acceptance A and fakerate F are slightly affected by the smearing, resulting in observable deviations only at small p_T^Z.

The systematic uncertainties are assumed uncorrelated, such that the total systematic uncertainty is given as the square root of the quadratic sum of all estimated sources. Figure 5.24 shows the systematic and the statistical uncertainties in several rapidity bins. Excluding the low-$p_{\mathrm{T}}^{\mathrm{Z}}$ region, where the JEC uncertainty supersedes the other uncertainty sources, the total systematic uncertainty of the triple differential measurement in $(\phi_{\eta}^{*}, y^{*}, y_{\mathrm{b}})$ is comparable to the one in $(p_{\mathrm{T}}^{\mathrm{Z}}, y^{*}, y_{\mathrm{b}})$. It ranges from less than 4% in the central rapidity region to 6% in the high-y_{b} and up to 20% in the high-y^{*} region.

The statistical uncertainty rarely exceeds the total systematic uncertainty across the entire phase space. Thus, the data collected in 2016 can be used for a precision analysis. However, the sparsely populated bins in the high-$p_{\mathrm{T}}^{\mathrm{Z}}$ or high-ϕ_{η}^{*} and the high-rapidity profit from the extension of the measurement to the full 13 TeV data taking period, corresponding to an integrated luminosity of $137\,\mathrm{fb}^{-1}$.

5.7 Comparison of cross section results to Monte Carlo simulations

The cross section results have been derived by normalizing the measured and unfolded distributions to the integrated luminosity of the dataset ($35.9\,\mathrm{fb}^{-1}$). They have been compared to the simulations created with the PYTHIA8 event generator. These simulations were created using matrix elements provided at LO precision by MADGRAPH5 for the P8+MG simulation and at NLO precision by MADGRAPH5_aMC@NLO for the P8+aMC simulation, as introduced in Section 5.3.2. Parton shower and hadronization modelling was performed by the event generator.

Figure 5.25 shows the results for the high-y^{*}, the high-y_{b} and the central rapidity bin. The statistical and systematic uncertainties of the measurements and the statistical uncertainties of the simulations are shown in the ratios. Since the statistical uncertainties of the given simulations, especially P8+aMC, increase towards high $p_{\mathrm{T}}^{\mathrm{Z}}$ or ϕ_{η}^{*} and the edge rapidity regions, large fluctuations can be observed.

Differences between the data and the P8+MG simulation beyond the uncertainties can be observed in most of the phase space. Especially towards high y^{*}, this simulation at LO precision does not yield a reasonable description of the data.

The deviations are significantly reduced when comparing the data to the P8+aMC simulation. This simulation at NLO precision is in good agreement with the data at low $p_{\mathrm{T}}^{\mathrm{Z}}$ or ϕ_{η}^{*}, even at high y^{*}. Small deviations remain in the medium range of $p_{\mathrm{T}}^{\mathrm{Z}}$ or ϕ_{η}^{*} in the low-y_{b} region and at low $p_{\mathrm{T}}^{\mathrm{Z}}$ or ϕ_{η}^{*} in the high-y_{b} region.

It is possible that the observed deviations can be covered by further uncertainties associated with the simulations, such as scale uncertainties, which are not accessible within the given simulations. Furthermore, as the large k-factors between the LO and the NLO simulation in the high-y^{*} region indicate, theory predictions at NNLO precision are required. The following Section 5.8 presents and discusses calculations which enable the estimation of theoretical uncertainties and provide NNLO precision in QCD.

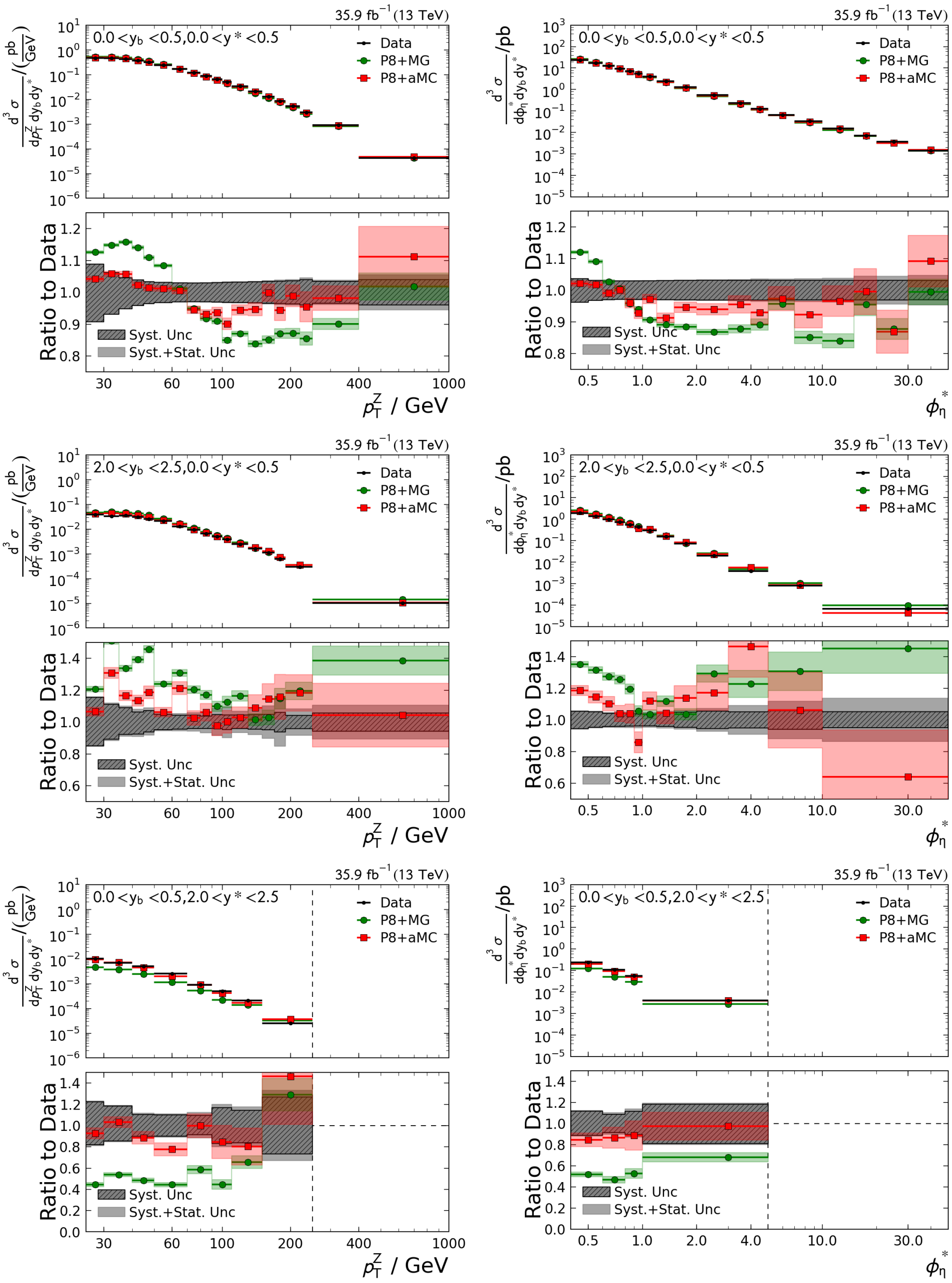


Figure 5.25: Comparison of measured cross sections to simulations generated with the Pythia8 event generator at LO (P8+MG) and NLO (P8+aMC) precision as a function of p_T^Z (left) and ϕ^*_η (right) in the central rapidity (top), the high-y_b (centre) and the high-y^* bin (bottom). In the ratios, the coloured bands represent the statistical uncertainties of the simulations (green and red). The total systematic uncertainty of the measurement is shown as hatched grey band, the quadratic sum of systematic and statistical uncertainties of the data is illustrated as light grey band. An overview of all rapidity bins can be found in Figure A.11 and A.12.

5.8 Theoretical predictions

The theoretical predictions for Z + jet production were calculated [71] and have been compared to the measured cross sections. These calculations were derived up to NNLO precision in QCD using an implementation by [7] which makes use of the NNLOJET parton-level Monte Carlo generator [18] and interpolation techniques provided by FASTNLO [15] to enable a flexible use of different PDFs and the variation of the renormalization and factorization scales.

Remark: The implementation was developed as part of a master's thesis [7] in order to automate the calculation of theory predictions at NNLO precision, which require an enormous amount of computing power. As part of a bachelor's thesis [72], the predictions presented within this thesis have been provided at a preliminary stage. In particular, a limitation in the number of available events can be observed as fluctuations.

Figure 5.26 presents the fixed order calculations up to NNLO in several rapidity bins, using PDFs from the NNPDF3.1 fit [3]. The scale was chosen as

$$E_{\mathrm{T}}^{\mathrm{Z}} = \sqrt{(m_{\mathrm{T}}^{\mathrm{Z}})^2 + (p_{\mathrm{T}}^{\mathrm{Z}})^2} \tag{5.16}$$

and the scale uncertainties have been derived by varying the renormalization and factorization scale, respectively, with a factor 0.5, 1.0 or 2.0. The scale uncertainty is taken as the maximum deviation of all possible combinations of these factors from the central result.

The ratio NLO/LO illustrates the k-factor, which takes values of up to 1.5 in the high-y_{b} and the central rapidity bin. It can be noted that the NNLO prediction agrees with the NLO prediction within their scale uncertainties, indicating a good convergence of the perturbative series in the low-y^* region.

Towards high y^*, the differences between the fixed order calculations increase, reaching NLO/LO k-factors of up to 14. The prediction at NNLO precision exhibits significantly reduced scale uncertainties and is therefore essential for a comparison to measurements in this phase space region.

The reason for the large NLO correction is that the parton-parton scattering subprocess composition differs strongly between the fixed order calculations, as discussed in Section 5.1. The fractional subprocess contributions to the respective cross section are presented in Figure 5.27 at LO, in Figure 5.28 at NLO and in Figure 5.29 at NNLO precision.

At LO precision, quark-antiquark, gluon-antiquark and quark-gluon scattering are the only possibilities for the creation of a Z + jet event, the additional processes of quark-quark, antiquark-antiquark and gluon-gluon scattering contribute only at higher orders. The dominant contribution in all regions of the phase space is quark-gluon scattering. Beyond LO, subprocesses can obtain negative weights as a result of interference between terms of the perturbative series. Nevertheless, the physical cross sections remain positive, such that the decompositions shown in the Figures 5.27, 5.28 and 5.29 always sum up to unity.

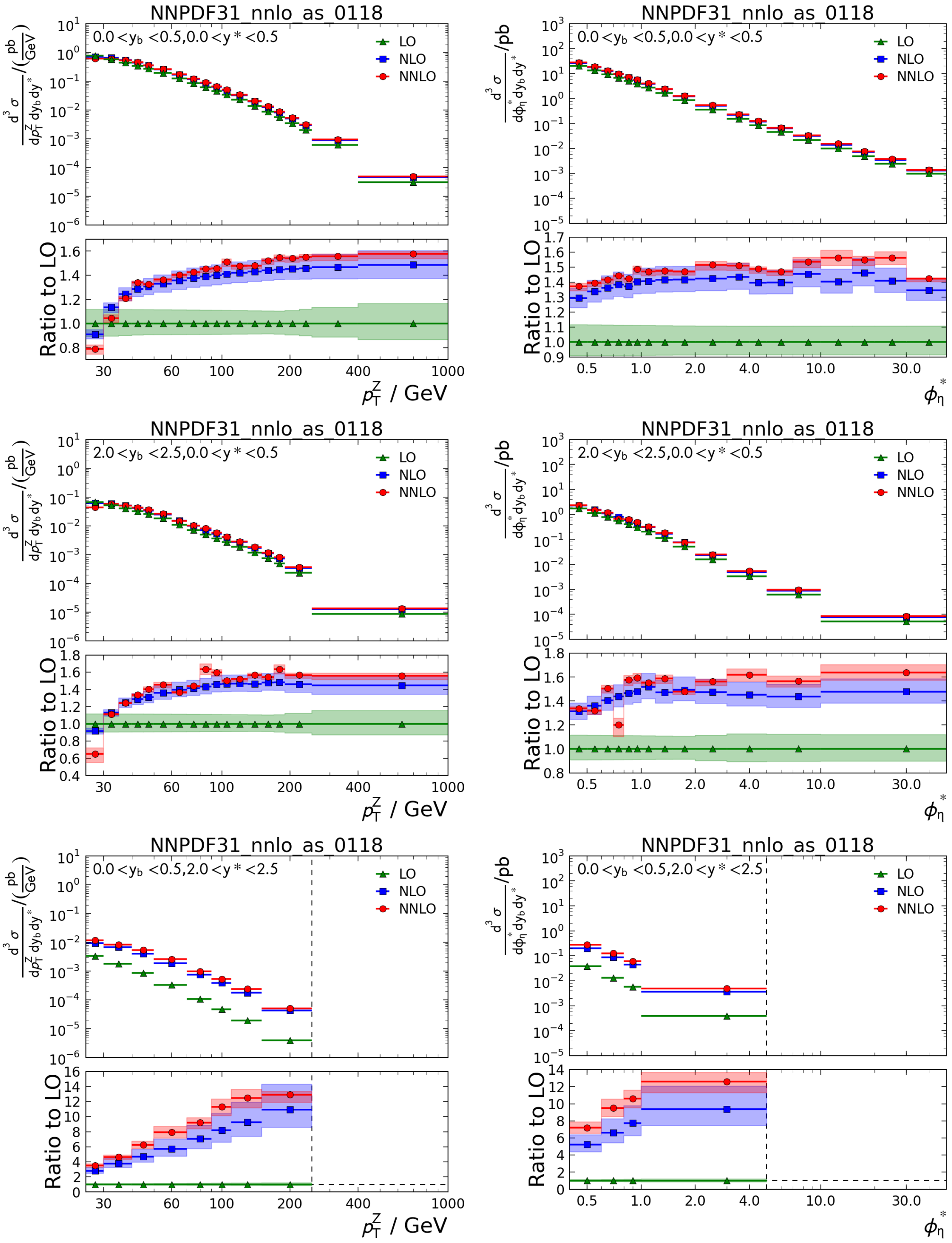


Figure 5.26: Comparison of distributions in LO, NLO and NNLO calculations as a function of $p_{\mathrm{T}}^{\mathrm{Z}}$ (left) and ϕ_{η}^{*} (right) in the central rapidity (top), the high-y_{b} (centre) and the high-y^{*} bin (bottom). The coloured bands indicate the scale uncertainty. An overview of all rapidity bins can be found in Figure A.13 and A.14.

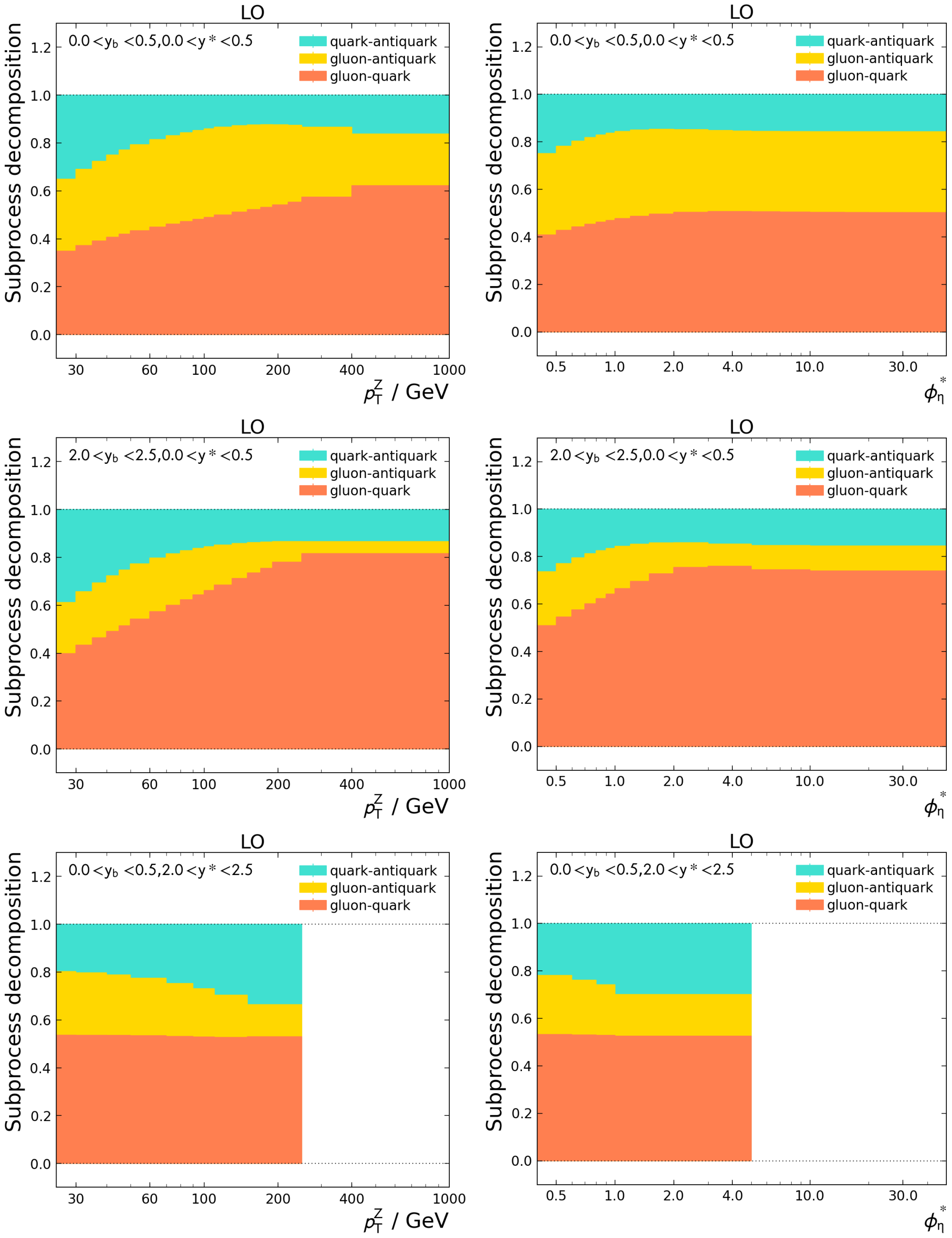


Figure 5.27: Subprocess decomposition of cross section calculation at LO as a function of $p_{\mathrm{T}}^{\mathrm{Z}}$ (left) and ϕ_{η}^{*} (right) in the central rapidity (top), the high-y_{b} (centre) and the high-y^{*} bin (bottom). An overview of all rapidity bins can be found in Figure A.15 and A.16.

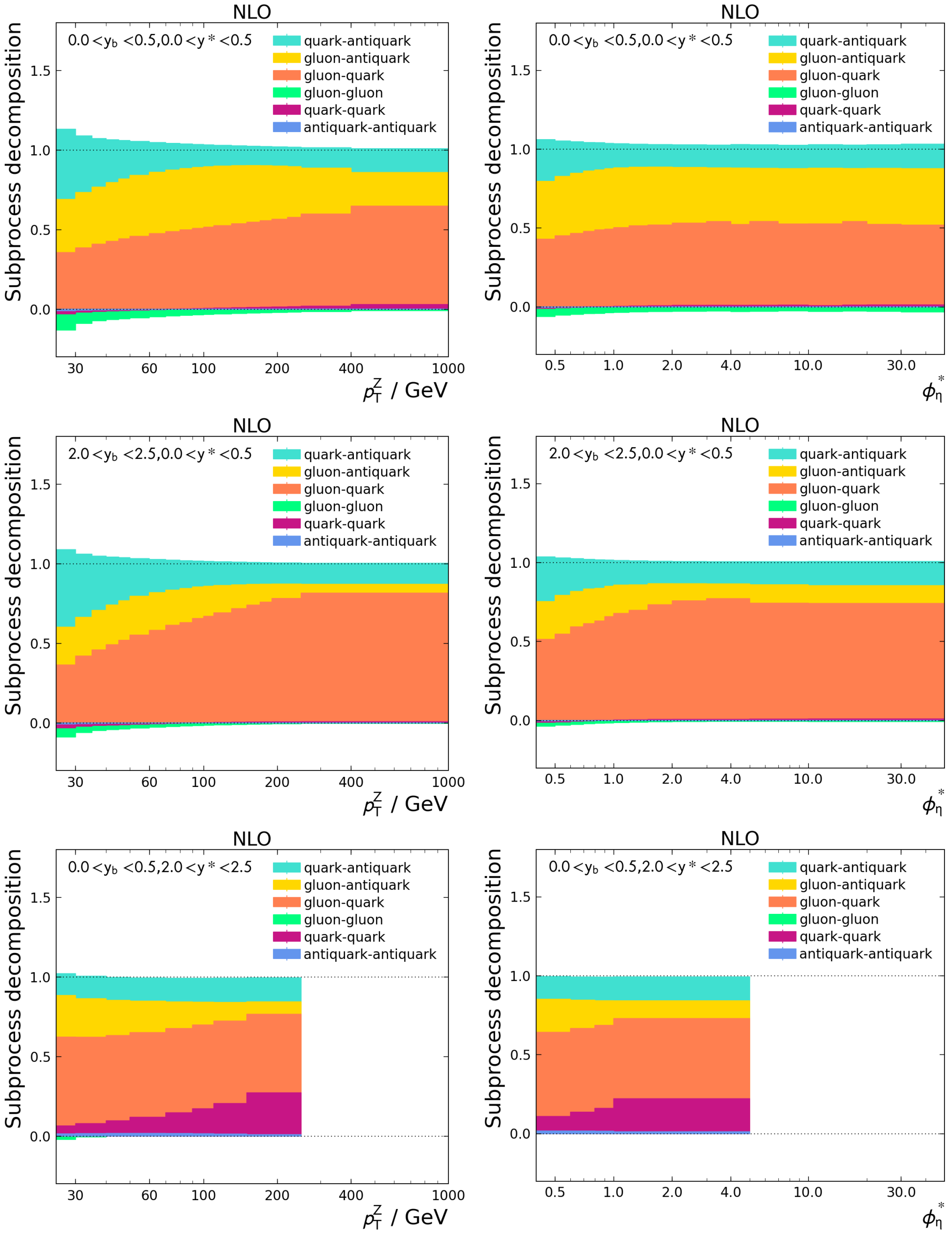


Figure 5.28: Subprocess decomposition of cross section calculation at NLO as a function of $p_{\mathrm{T}}^{\mathrm{Z}}$ (left) and ϕ^*_η (right) in the central rapidity (top), the high-y_{b} (centre) and the high-y^* bin (bottom). Subprocesses can contribute with negative signs at NLO. An overview of all rapidity bins can be found in Figure A.17 and A.18.

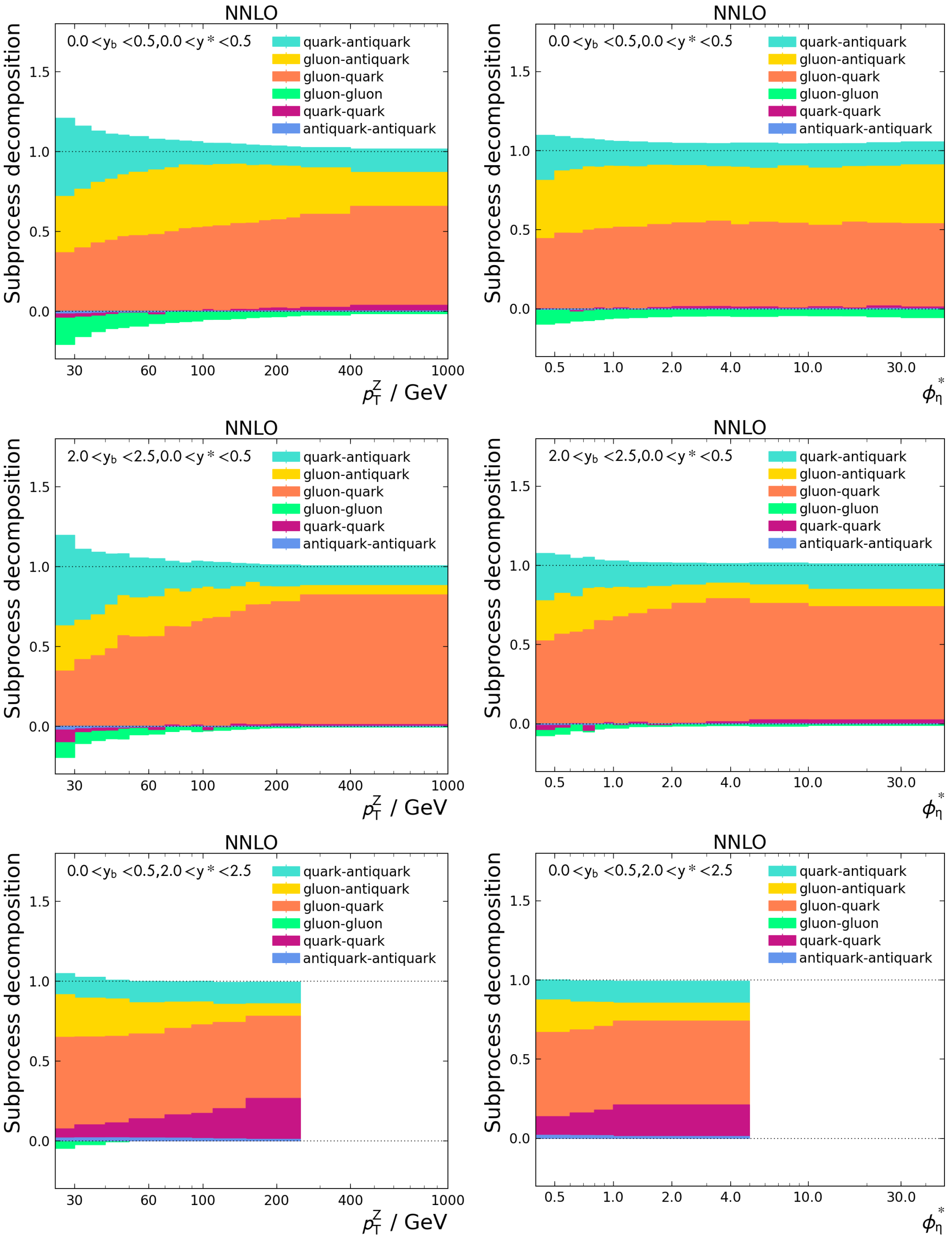


Figure 5.29: Subprocess decomposition of cross section calculation at NNLO as a function of p_T^Z (left) and ϕ^*_η (right) in the central rapidity (top), the high-y_b (centre) and the high-y^* bin (bottom). Subprocesses can contribute with negative signs at NNLO. An overview of all rapidity bins can be found in Figure A.19 and A.20.

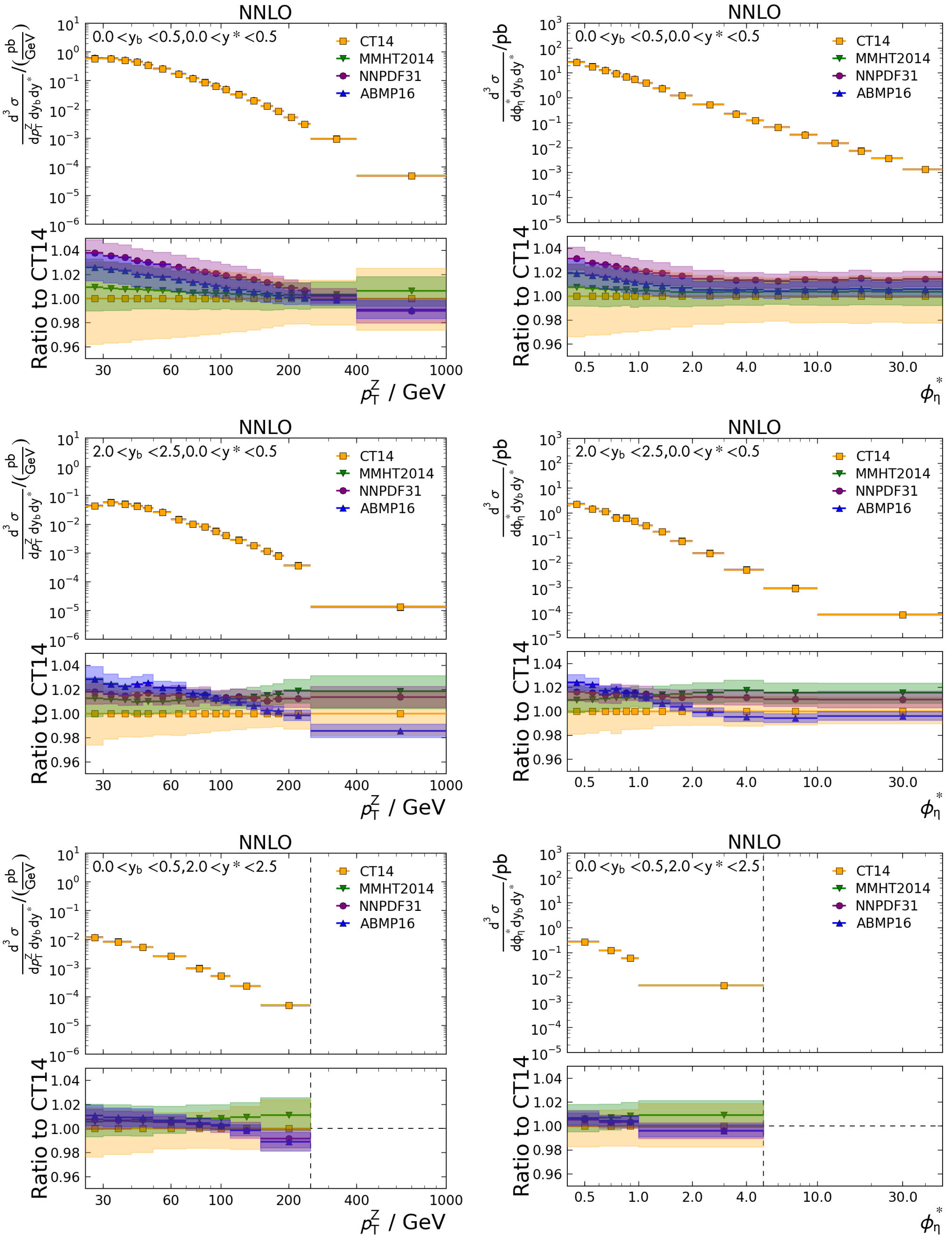


Figure 5.30: Comparison of distributions calculated at NNLO for different PDF sets as a function of p_T^Z (left) and ϕ_η^* (right) in the central rapidity (top), the high-y_b (centre) and the high-y^* bin (bottom). The coloured bands indicate the PDF uncertainty. An overview of all rapidity bins can be found in Figure A.21 and A.22.

The relation between the momentum fractions of the colliding partons is preserved in y^* and y_b. Large values of y^*, associated with small values of y_b, require momentum fractions of the same magnitude. While an up or a down quark in a proton tends to carry a large momentum fraction, an antiquark or a gluon tends to carry a comparatively low momentum fraction. Thus large values of y^* are unlikely to occur unless higher orders are included, resulting in large NLO corrections.

Figure 5.30 shows the calculated distributions at NNLO using the fit results of other PDF studies besides NNPDF3.1: CT14 [4], MMHT2014 [5] and ABMP16 [6]. The differences are in the range of a few percent.

5.9 Comparison of cross section results to fixed order calculations

The measured cross section results, normalized to the integrated luminosity of the dataset ($35.9\,\mathrm{fb}^{-1}$), are compared to the theory predictions at NLO and NNLO precision described in the previous Section 5.8, using PDFs from the NNPDF3.1 fit.

Figure 5.31 shows the results for the high-y^*, the high-y_b and the central rapidity bins. The statistical and systematic uncertainties of the measurements are shown in the ratios, as well as the theory uncertainties, determined as the quadratic sum of the scale and the PDF uncertainties.

In most phase space regions, the observed differences between the measurements and the NLO calculations are compatible with the theory uncertainties. However, in the low-y^* regions, deviations of up to 25% can be observed at low p_T^Z. These phase space regions are known to be sensitive to additional effects which are not included in the calculations, such as non-perturbative corrections or parton showering. The inclusion of these effects is expected to reduce this discrepancy.

At NNLO precision, deviations of up to 20%, which are not covered by the uncertainties, are exposed in the high-y_b region. Since y_b contains information about the kinematics of the colliding partons, any discrepancy in this region hints to a sensitivity to the PDFs. The subprocess decomposition in this phase space region is almost exclusively determined by quark-gluon scattering, making it particularly sensitive to the gluon PDF.

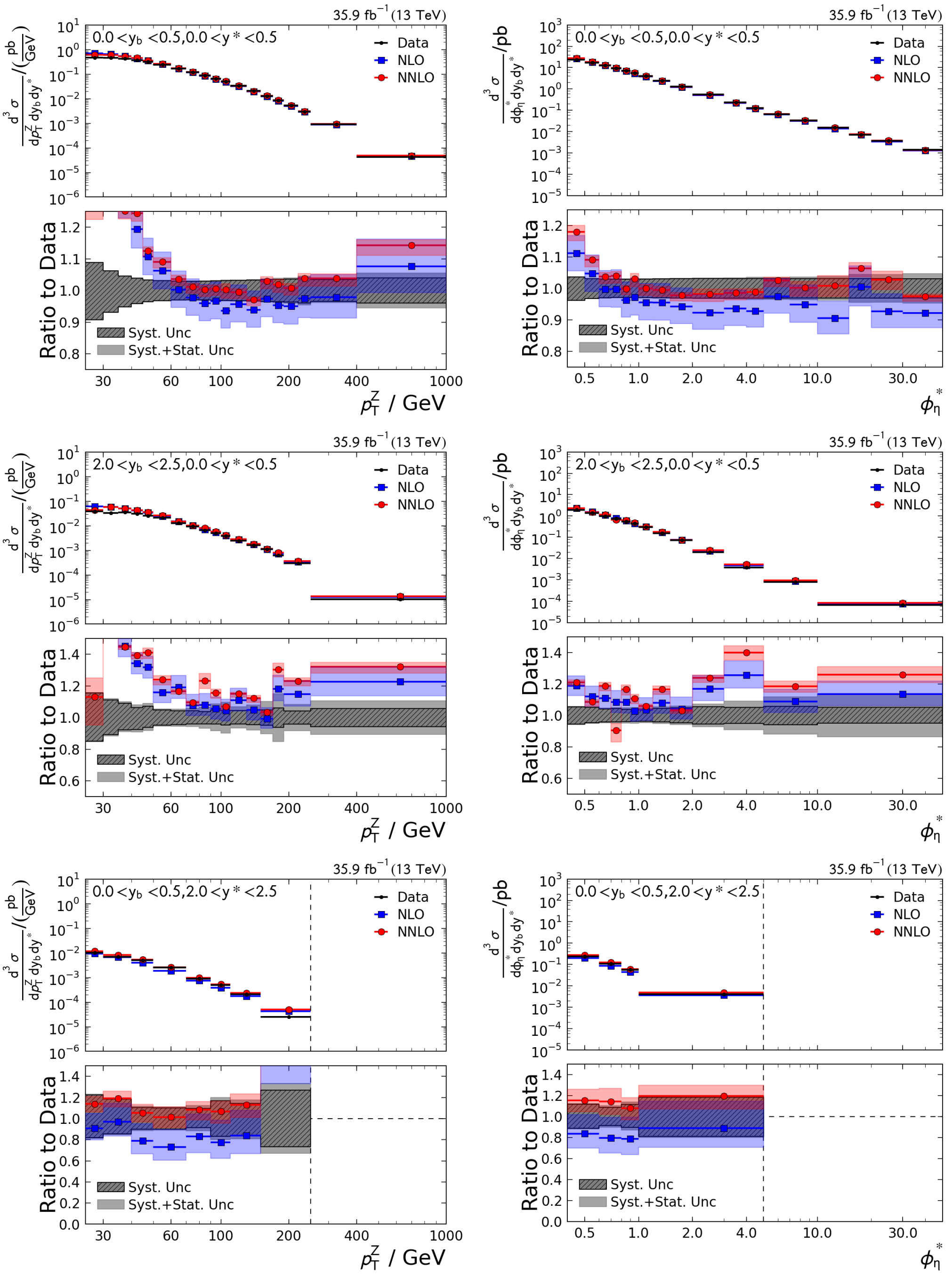


Figure 5.31: Comparison of measured cross sections to fixed order calculations at NLO and NNLO precision as a function of $p_{\mathrm{T}}^{\mathrm{Z}}$ (left) and ϕ^*_η (right) in the central rapidity (top), the high-y_{b} (centre) and the high-y^* bin (bottom). In the ratios, the coloured bands represent the theory uncertainties of the calculations (blue and red). The total systematic uncertainty of the measurement is shown as hatched grey band, the quadratic sum of systematic and statistical uncertainties of the data is illustrated as light grey band. An overview of all rapidity bins can be found in Figure A.23 and A.24.

5.10 Summary

In this chapter, the first measurement of the triple differential inclusive Z ($\rightarrow \mu\mu$) + jet cross section at a centre-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$ has been performed and presented as a function of ($p_{\mathrm{T}}^{\mathrm{Z}}, y^{*}, y_{\mathrm{b}}$) and ($\phi_{\eta}^{*}, y^{*}, y_{\mathrm{b}}$).

Data recorded by the CMS detector in the year 2016 have been used, corresponding to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$. The cross sections have been corrected for detector effects using three-dimensional unfolding based on forward smearing.

The results have been compared to Monte Carlo simulations at LO and NLO precision and to preliminary fixed order calculations at NLO and NNLO precision. The measurements have been found in good agreement with the predictions over a wide phase space range. Differences between the fixed order calculations and the measurements beyond experimental and theoretical uncertainties have been identified at the low-$p_{\mathrm{T}}^{\mathrm{Z}}$ and the high-y_{b} region, with the latter exposed only at NNLO precision.

In the low-$p_{\mathrm{T}}^{\mathrm{Z}}$ region, additional effects such as non-perturbative corrections or parton showering become relevant. Corrections concerning these effects which are missing in the theory calculations can be included in the analysis once they have been provided.

In the high-y_{b} region, the quark-gluon dominance of the subprocess decomposition increases. Limited knowledge of the gluon PDF can therefore explain the deviations observed in this region. Conversely, the inclusion of the results of this thesis in PDF fits is expected to be used to further improve the gluon PDF.

CHAPTER 6

Conclusion

The goal of this thesis was the analysis of Z boson production in association with jets based on data recorded with the CMS detector at the LHC. The data correspond to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$ of proton-proton collisions at a centre-of-mass energy of 13 TeV. Z bosons decaying to a muon-antimuon pair are of particular interest for tasks that require high precision, of which two were performed and presented within this thesis.

The analysis of Z ($\rightarrow \mu\mu$) + jet events plays an important role in the process of jet energy calibration, which has been presented in Chapter 4. To account for residual differences between data and simulation regarding the jet reconstruction in the CMS detector, the transverse momenta of jets are calibrated in reference to Z bosons. Two complementary methods have been used within this thesis to derive jet response values on an event-by-event basis in both data and simulation. Whereas the p_{T} balance method, based on the ratio between the transverse momenta of the jet and the Z boson, provides a straightforward estimation of jet miscalibration, the missing transverse momentum projection fraction (MPF) method, based on the missing transverse momentum in the event, takes the entire detector configuration into account. There are certain advantages and challenges for each method, one of which is their differing sensitivity to additional jet activity. The determined response values have been extrapolated to an exclusive Z ($\rightarrow \mu\mu$) + jet topology to confirm the consistency of both methods.

The data-to-simulation ratio of the responses is used to perform a global fit, combining results from several working groups within the CMS collaboration. Different reference objects are used to maximize the accuracy of the obtained residual corrections. The Z ($\rightarrow \mu\mu$) + jet calibration channel providing the highest precision forms a crucial input for the fit procedure.

The resulting jet energy corrections contribute to most physics analyses which take the 2016 dataset into account. Regarding the challenges of the data taking in 2016, the collaboration achieved an outstanding performance of jet reconstruction with the CMS detector.

The discovery of additional detector issues during the 2017 and 2018 calibration necessitates further ongoing studies on their effects in the 2016 jet energy calibration. Therefore, the publication of the final results of the jet energy calibration for the complete 13 TeV dataset is expected within the next years, which is beyond the scope of this thesis.

The centre-of-mass energy and the increasing integrated luminosities available at the LHC give access to sensitivity in phase space regions which have not been accessible in the past. In Chapter 5, the first triple differential cross section measurement of the inclusive Z ($\rightarrow \mu\mu$) + jet process has been presented. The measurement has been performed as a function of the rapidity variables y^* and y_b, and the transverse momentum of the Z boson p_T^Z or the variable ϕ_η^*. The variable ϕ_η^* is an observable derived from muon angular information which is correlated to p_T^Z but can be measured more precisely.

The use of y^* given as the rapidity separation between the Z boson and the leading jet, and y_b, given as the boost of their centre-of-mass system, provides a suitable division of the rapidity phase space. The scattering angle corresponds to y^* and carries information about the matrix element, the boost y_b carries information about the kinematics of the colliding partons determined by their parton distribution functions (PDFs). Thus, effects originating from the matrix elements can be disentangled from the PDFs, such that an increased PDF sensitivity can be reached.

The cross sections measured as a function of $(p_\mathrm{T}^\mathrm{Z}, y^*, y_\mathrm{b})$ or $(\phi_\eta^*, y^*, y_\mathrm{b})$ were corrected for detector effects by unfolding. Since the available simulations were not optimized for a three-dimensional measurement approach, they exhibit insufficient statistical significance in certain phase space regions. Some effort had to be made by choosing a suitable binning and implementing a forward smearing approach to create response matrices for unfolding. This procedure implied a number of assumptions and approximations which have been validated by closure tests. As a result, a sizeable reduction of the uncertainties associated with unfolding could be reached.

The total systematic uncertainties of the measurements have been estimated to range between 4 % in low ϕ_η^* and central rapidity regions and 20 % in the high y^* region. The statistical uncertainties provided by the 2016 dataset have been found to exceed the systematic uncertainties in high p_T^Z or high ϕ_η^* regions. These regions profit from the extension of the measurement to the full 13 TeV data taking period, corresponding to an integrated luminosity of 137 fb^{-1}.

In the high y^* region, the unfolding uncertainty, which itself is dominated by the difference between results derived from the Pythia and the Herwig++ event generators, respectively, has to be reduced when more data becomes available. Besides a higher event number in simulations, which facilitates the forward smearing procedure, a study of the tuning parameters which lead to the observed differences could be supportive to improve the measurement.

The measurement is summarized in Figure 6.1, which presents the measured cross sections compared to preliminary theory predictions calculated at NNLO precision, and in Figure 6.2, which presents the data-to-theory ratio as well as experimental and theoretical uncertainties. The measurements have been found to be in agreement with the predictions across a wide phase space range. Discrepancies observed at low p_T^Z are expected to be reduced by correcting the fixed order calculations for non-perturbative effects and parton shower contributions.

In comparison to predictions calculated at NLO precision, the calculations at NNLO precision have revealed differences to the measurements towards the high y_b region. The observable y_b is characterized by an exemplary sensitivity to the scattering subprocesses, so that any discrepancy can be traced back to PDFs. Because of the increased domination of the quark-gluon scattering process, including this measurement to PDF fits is expected to further improve the gluon PDF.

The measurement will be reprocessed once the jet energy calibration has been finalized and the missing corrections have been included into final NNLO calculations. Constraints on PDFs based on this measurement will be available and published subsequently.

In summary, this thesis has explored the possible benefits of high precision associated with the analysis of Z ($\rightarrow \mu\mu$) + jets events and confirmed that analyzing the full 13 TeV dataset can be expected to provide a significant contribution to PDF extraction. The results will have an impact on future measurements and calculations and thus, referencing to the famous quote at the beginning of the thesis, contribute to our understanding of what *"holds the world together in its inmost folds"*.

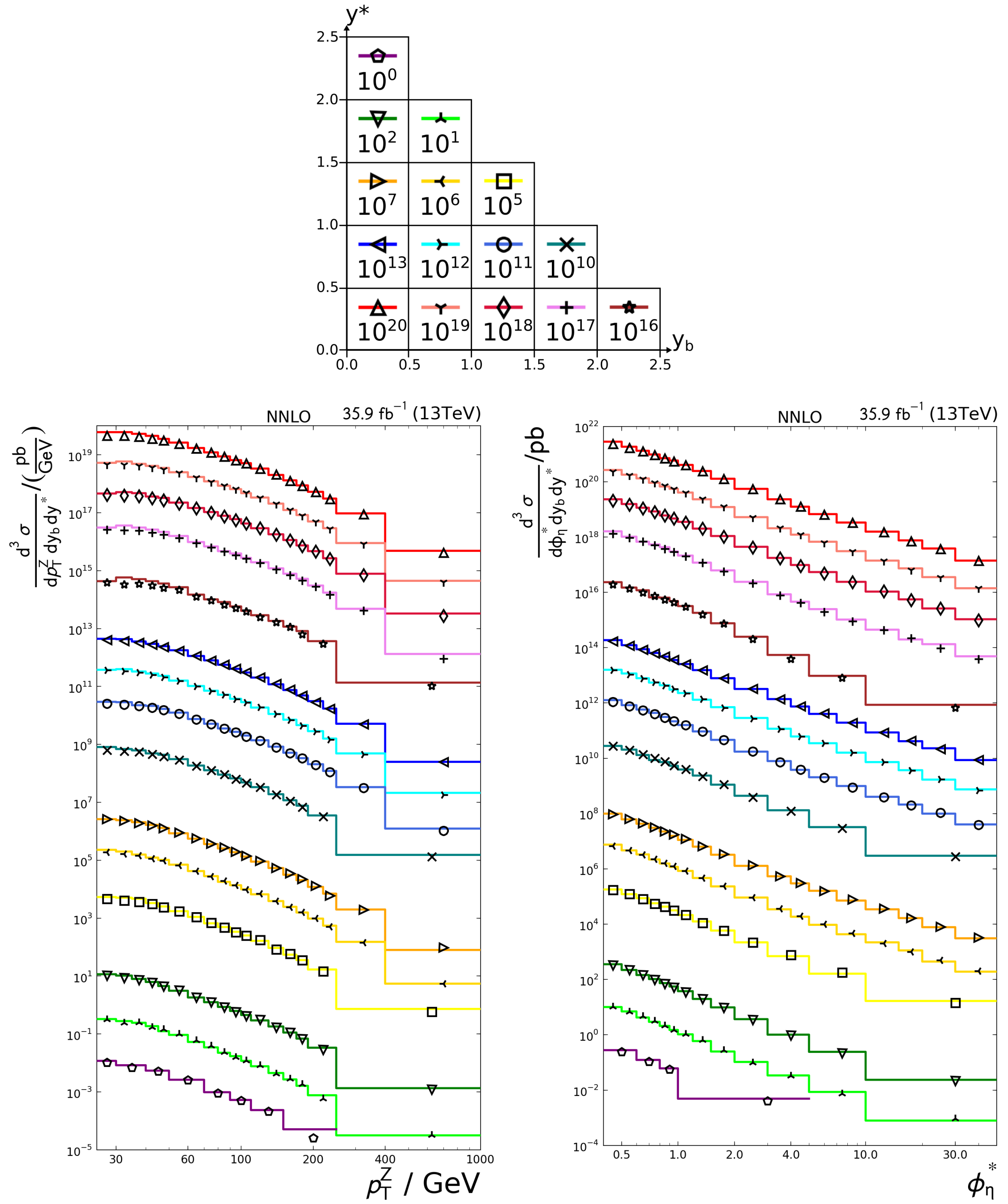


Figure 6.1: The triple differential inclusive Z ($\rightarrow \mu\mu$) + jet cross section measurements as a function of (p_T^Z, y^*, y_b) (left) and (ϕ^*_η, y^*, y_b) (right), shown as symbols and compared to preliminary theory predictions at NNLO precision shown as lines. The legend at the top illustrates which colour and symbol represent which rapidity bin, as well as which scaling factor is applied to improve readability.

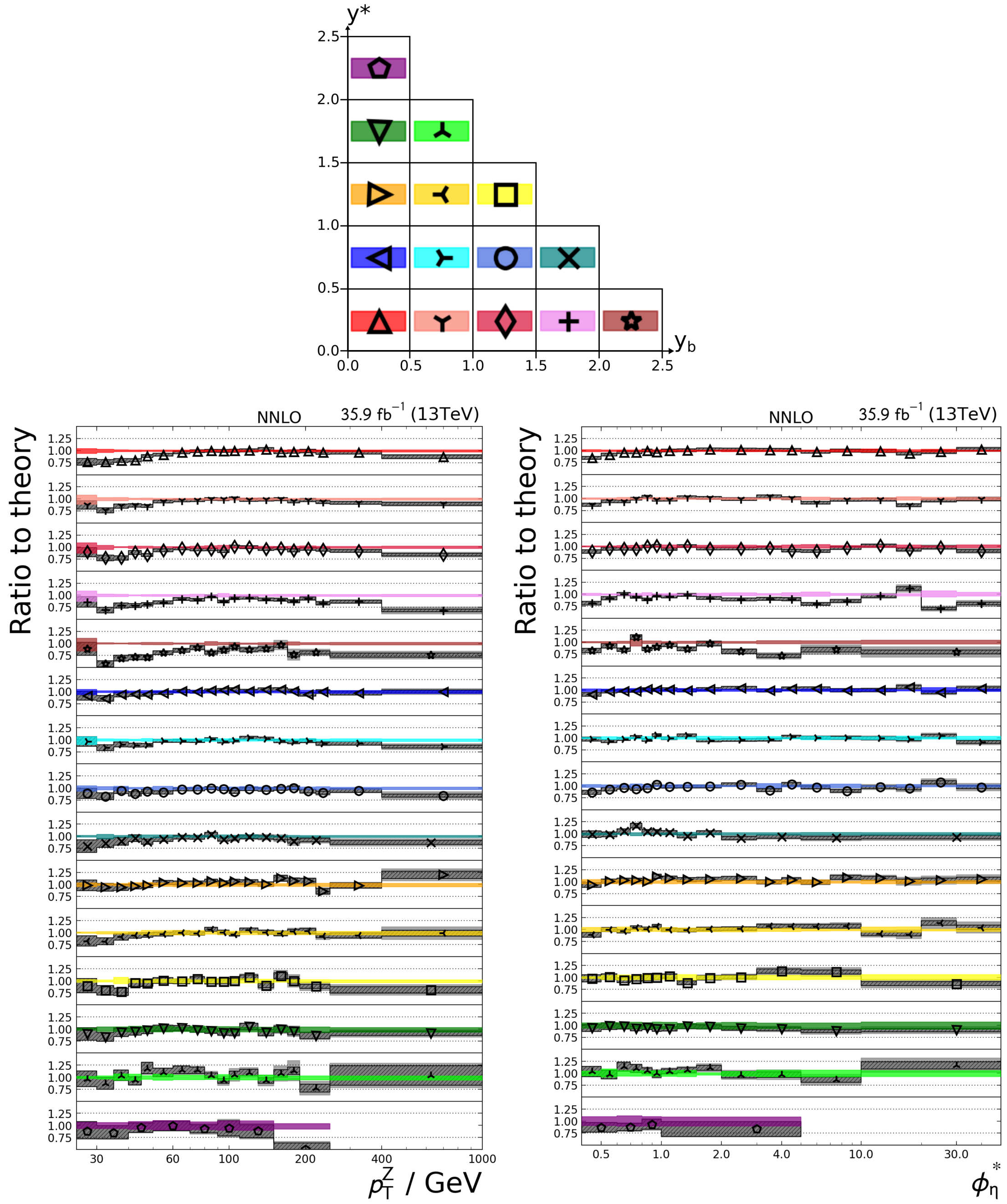


Figure 6.2: The ratios of the triple differential inclusive Z ($\rightarrow \mu\mu$) + jet cross section measurements to the preliminary theory predictions calculated at NNLO precision as a function of ($p_\mathrm{T}^\mathrm{Z}, y^*, y_\mathrm{b}$) (left) and ($\phi_\eta^*, y^*, y_\mathrm{b}$) (right), shown as symbols. The theory uncertainties are shown as coloured bands, the systematic uncertainties as hatched grey bands and the total experimental uncertainty, i.e. the quadratic sum of systematic and statistical uncertainties of the data as light grey bands. The legend at the top illustrates which colour and symbol represent which rapidity bin.

APPENDIX A

Appendix

Table A.1: Used datasets, the data certification file providing the information about certified luminosity sections and the integrated luminosity resulting from the data certification.

Datasets:	/DoubleMuon/Run2016*-17Jul2018*/MINIAOD
	/SingleMuon/Run2016*-17Jul2018*/MINIAOD
Data certification:	`Cert_271036-284044_`
	`13TeV_23Sep2016ReReco_Collisions16_JSON.txt`
Integrated luminosity:	$35.9\,\mathrm{fb}^{-1}$
Trigger flags:	`HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ`
	`HLT_IsoMu24` or `HLT_IsoTkMu24`

Table A.2: Simulated Monte Carlo simulations taken into account.

Calibration
/DY1JetToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[. . .]
/DY2JetToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[. . .]
/DY3JetToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[. . .]
/DY4JetToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[. . .]
Signal
/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[. . .]
/DYJetsToLL_M-50_TuneCUETHS1_13TeV-madgraphMLM-herwigpp/[. . .]
/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/[. . .]
Background
/ZZ_TuneCUETP8M1_13TeV-pythia8/[. . .]
/WZJToLLLNu_TuneCUETP8M1_13TeV-amcnlo-pythia8/[. . .]
/WWTo2L2Nu_13TeV-powheg/[. . .]
/TTJets_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[. . .]
/ST_tW_*top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/[. . .]
[. . .]
RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6*/MINIAODSIM

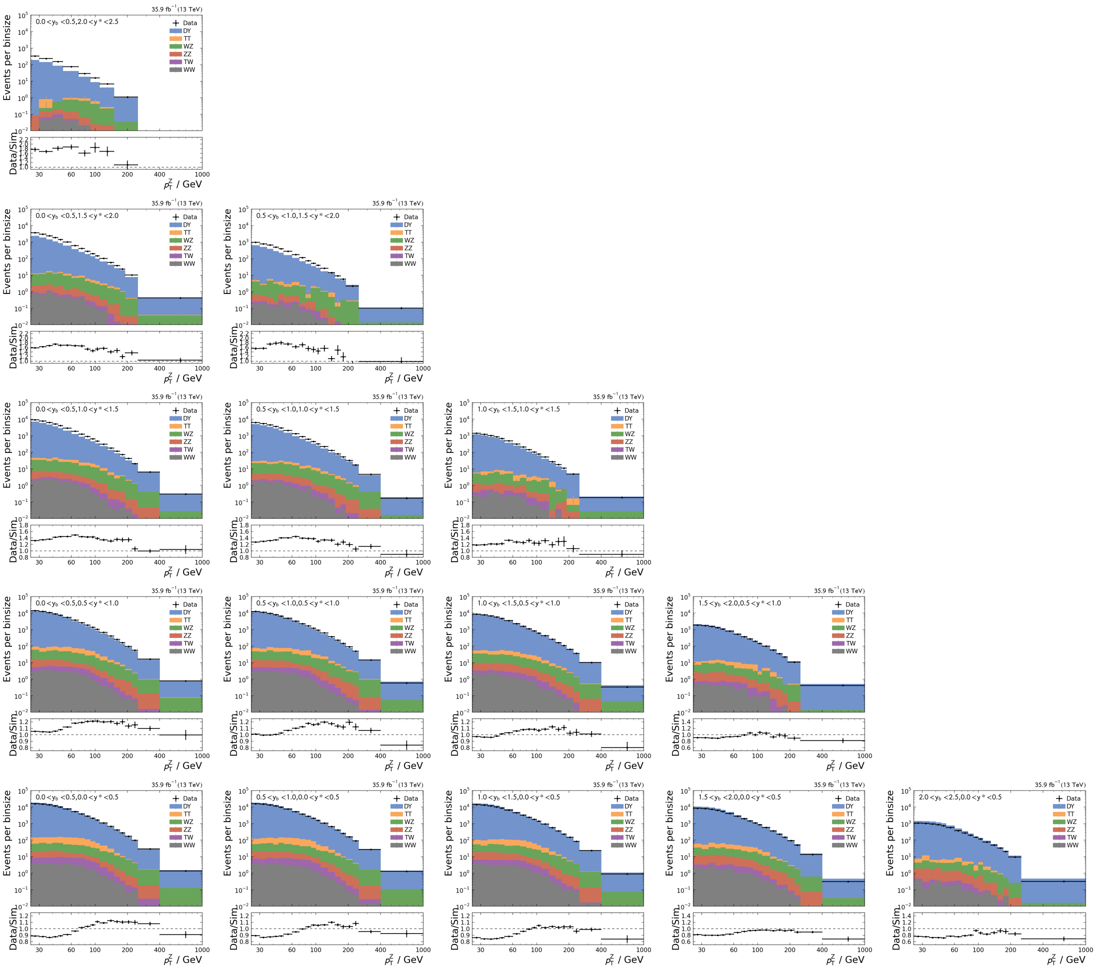


Figure A.1: The p_T^Z distribution in all rapidity bins for data and the P8+MG simulation and backgrounds. The simulations have been normalized to the luminosity of the data.

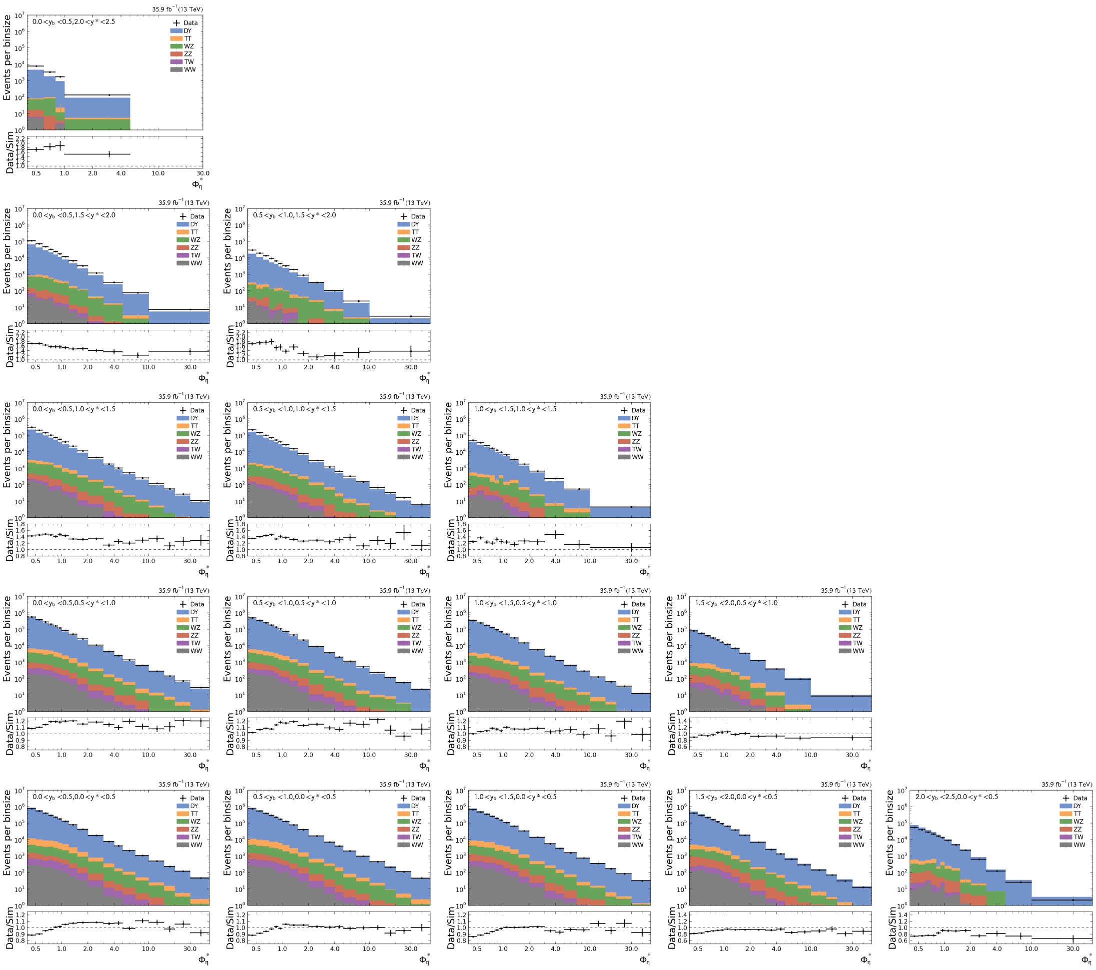


Figure A.2: The ϕ_η^* distribution in all rapidity bins for data and the P8+MG simulation and backgrounds. The simulations have been normalized to the luminosity of the data.

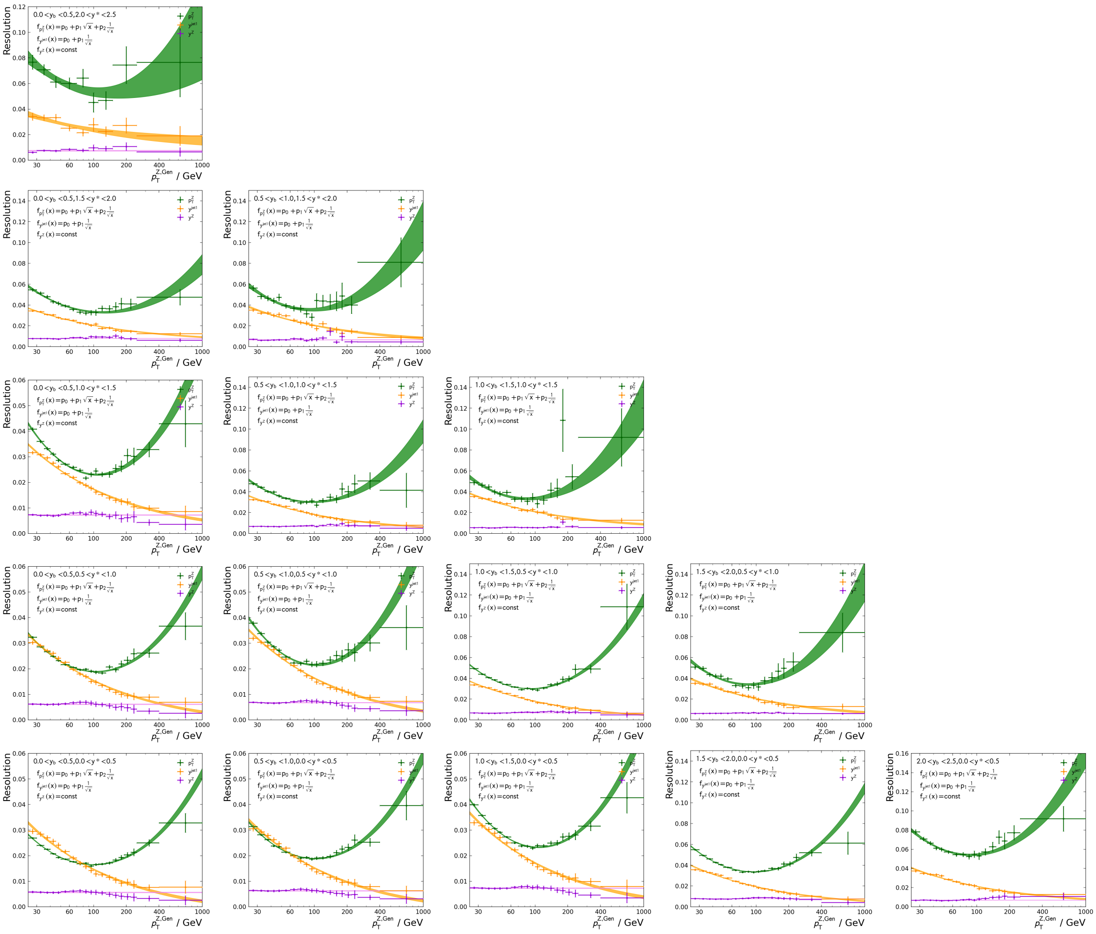


Figure A.3: The RMS results of histograms filled with $R(p_T^Z)$, $R(y^Z)$ and $R(y^{\mathrm{jet1}})$ in bins of p_T^Z in all rapidity bins. Different parametrizations have been chosen to smoothen the p_T^Z dependence. The RMS uncertainties are scaled to account for bias by the chosen parametrization and non-Gaussian behaviour of the distributions.

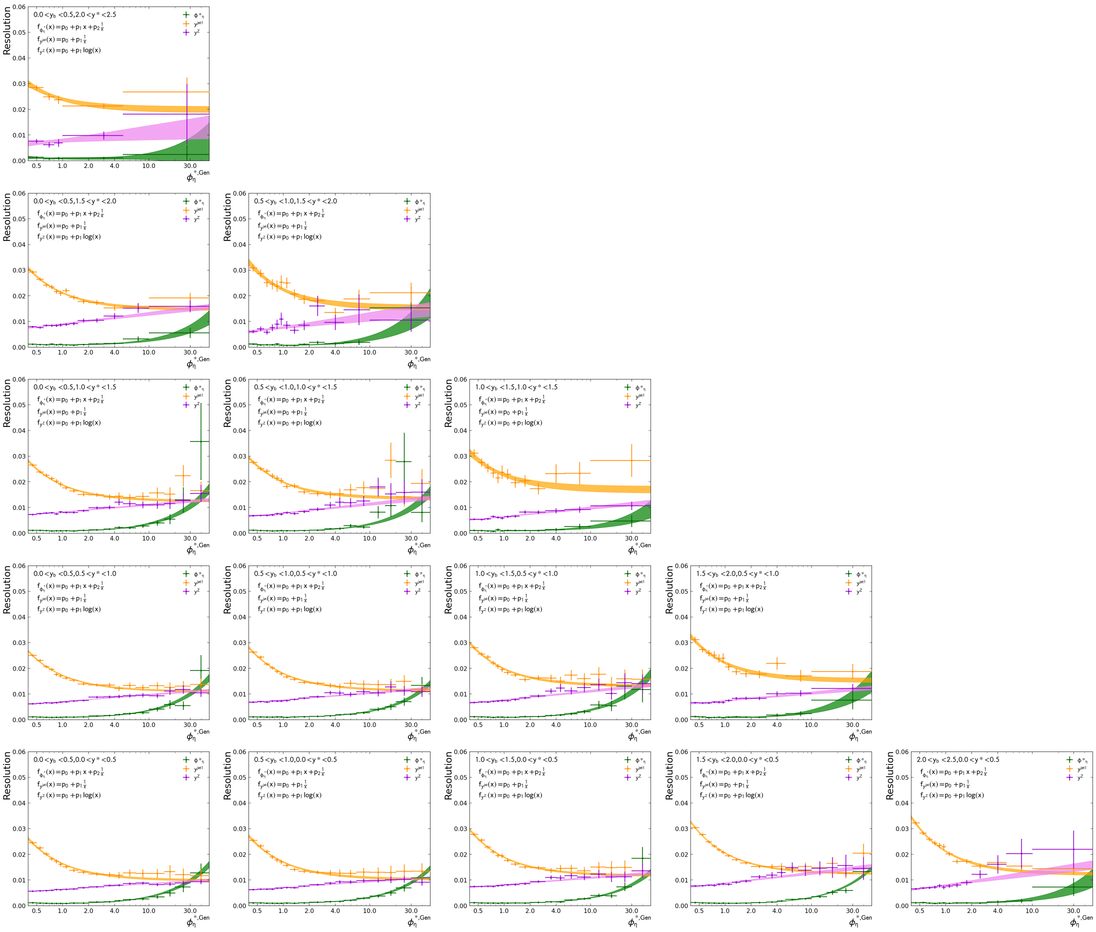


Figure A.4: The RMS results of histograms filled with $R(\phi^*_\eta)$, $R(y^\mathrm{Z})$ and $R(y^\mathrm{jet1})$ in bins of ϕ^*_η in all rapidity bins. Different parametrizations have been chosen to smoothen the ϕ^*_η dependence. The RMS uncertainties are scaled to account for bias by the chosen parametrization and non-Gaussian behaviour of the distributions.

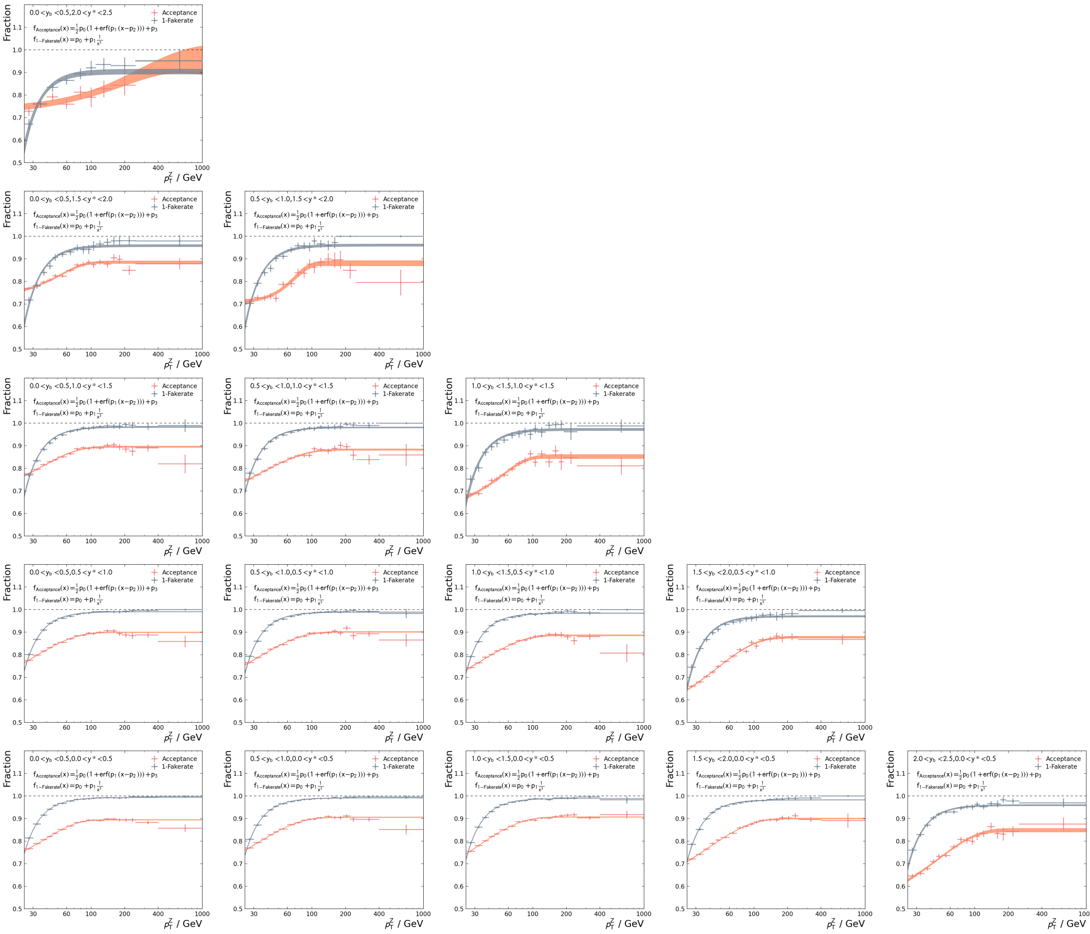


Figure A.5: Acceptance and fakerate in bins of $p_{\mathrm{T}}^{\mathrm{Z}}$ in all rapidity bins. Parametrizations have been chosen to smoothen the $p_{\mathrm{T}}^{\mathrm{Z}}$ dependence. The binomial uncertainties are scaled to account for bias by the chosen parametrization and deviations from the binomial approximation.

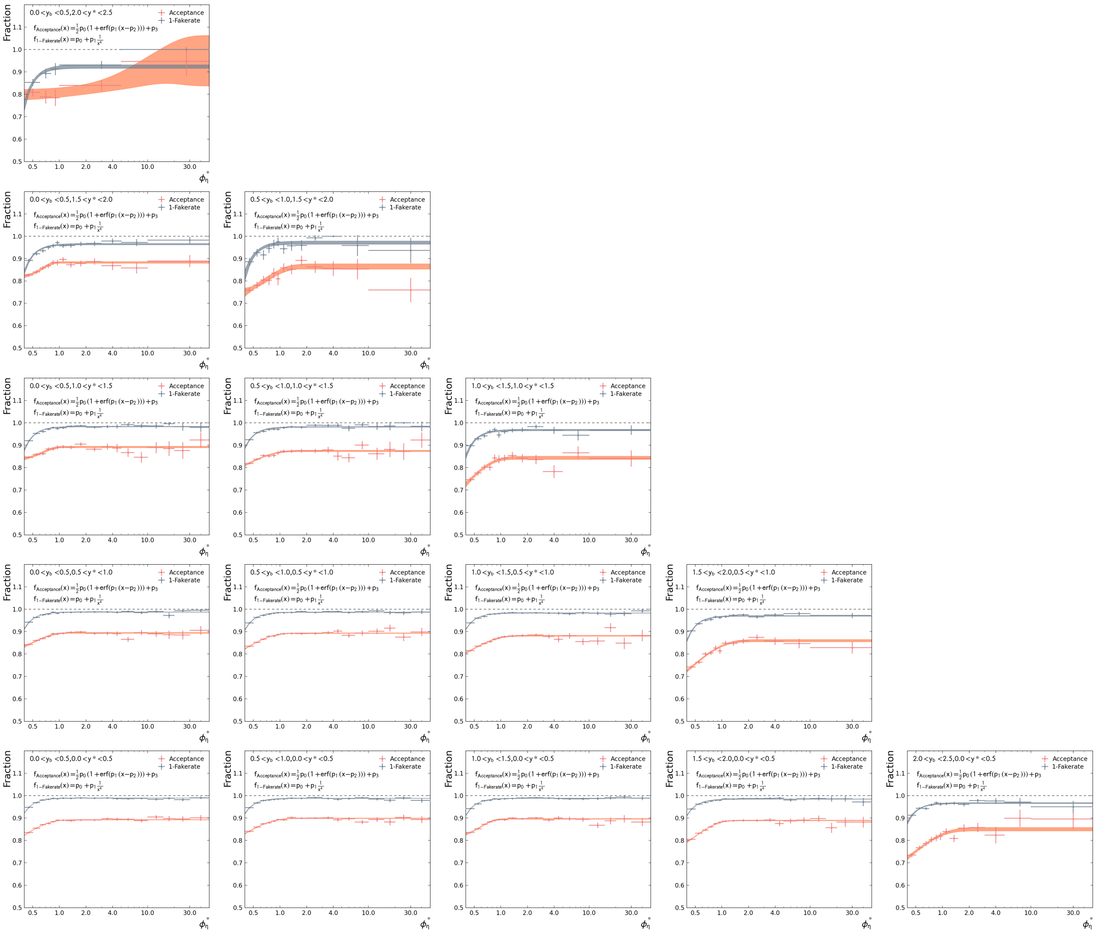


Figure A.6: Acceptance and fakerate in bins of ϕ^*_η in the all rapidity bins. Parametrizations have been chosen to smoothen the ϕ^*_η dependence. The binomial uncertainties are scaled to account for bias by the chosen parametrization and deviations from the binomial approximation.

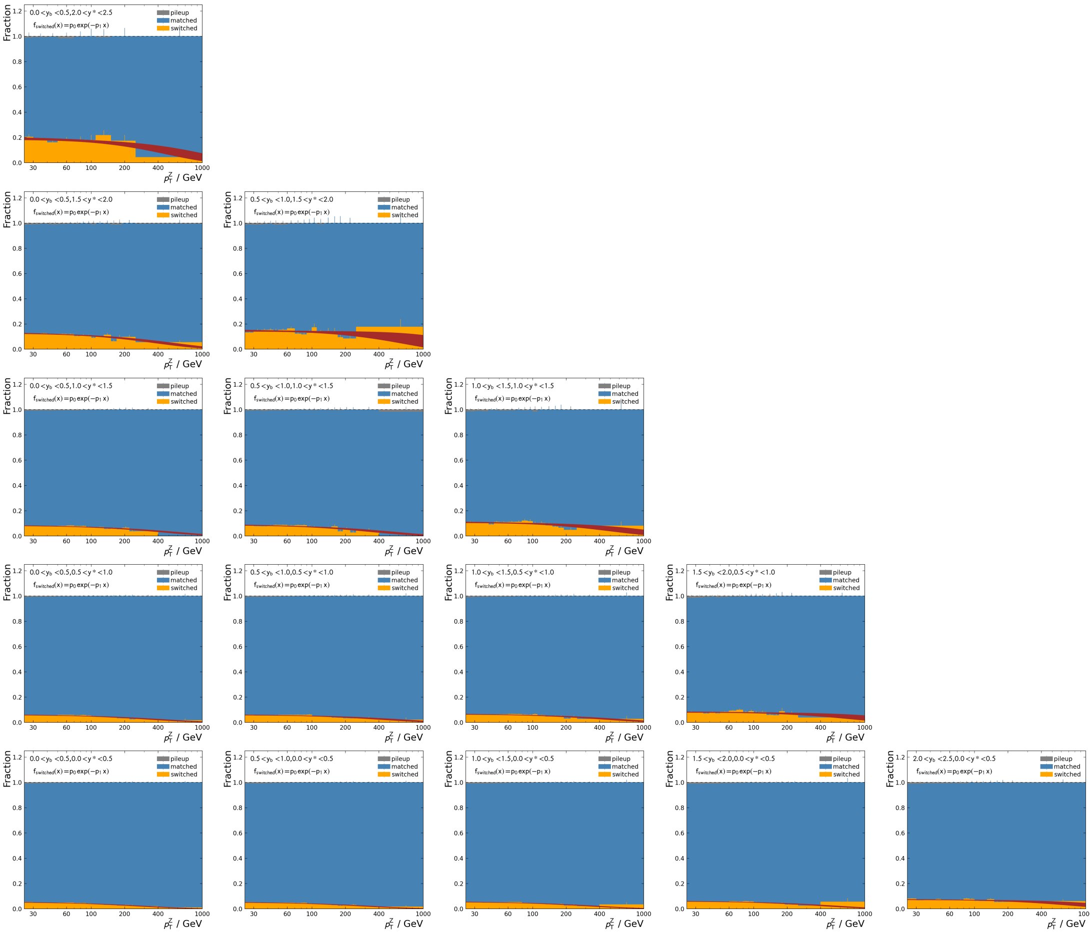


Figure A.7: Fraction of matched, switched and pileup events in bins of $p_{\mathrm{T}}^{\mathrm{Z}}$ in all rapidity bins. The switching fraction is parametrized by a suitable function. Its binomial uncertainties are scaled to account for bias by the chosen parametrization and deviations from the binomial approximation.

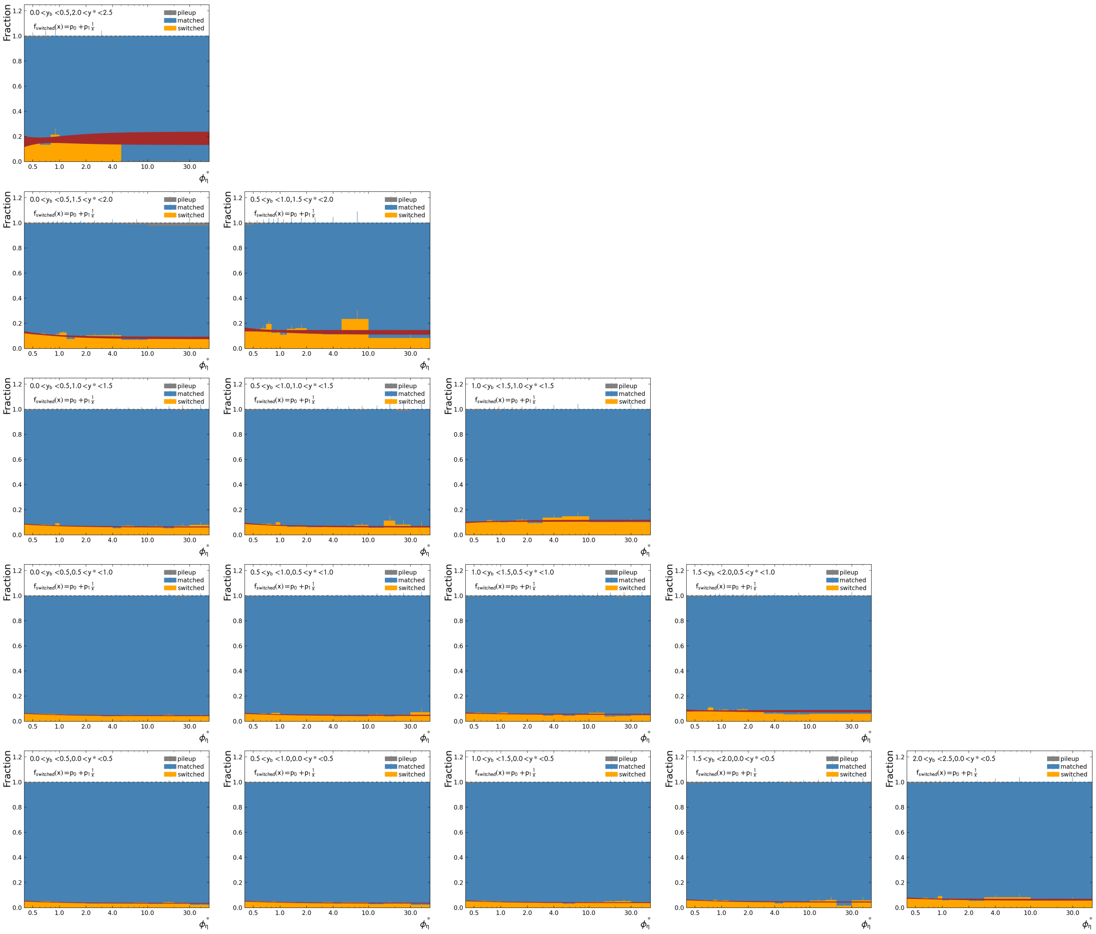


Figure A.8: Fraction of matched, switched and pileup events in bins of ϕ^*_η in all rapidity bins. The switching fraction is parametrized by a suitable function. Its binomial uncertainties are scaled to account for bias by the chosen parametrization and deviations from the binomial approximation.

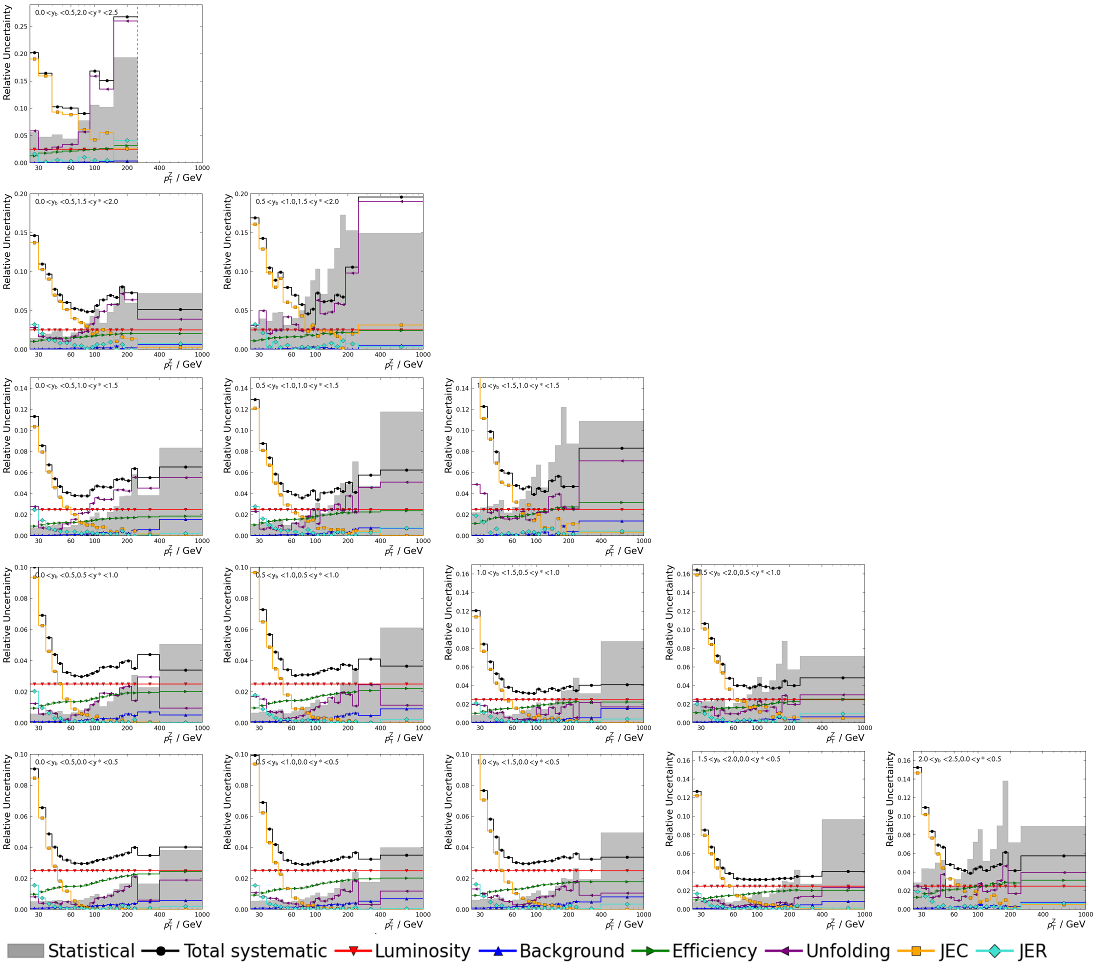


Figure A.9: Total systematic uncertainty (black) and the different uncertainty sources (coloured) as a function of $p_{\mathrm{T}}^{\mathrm{Z}}$ in all rapidity bins, compared to the statistical uncertainty (light grey). The uncertainties have been symmetrized by taking half of the spread in each uncertainty source.

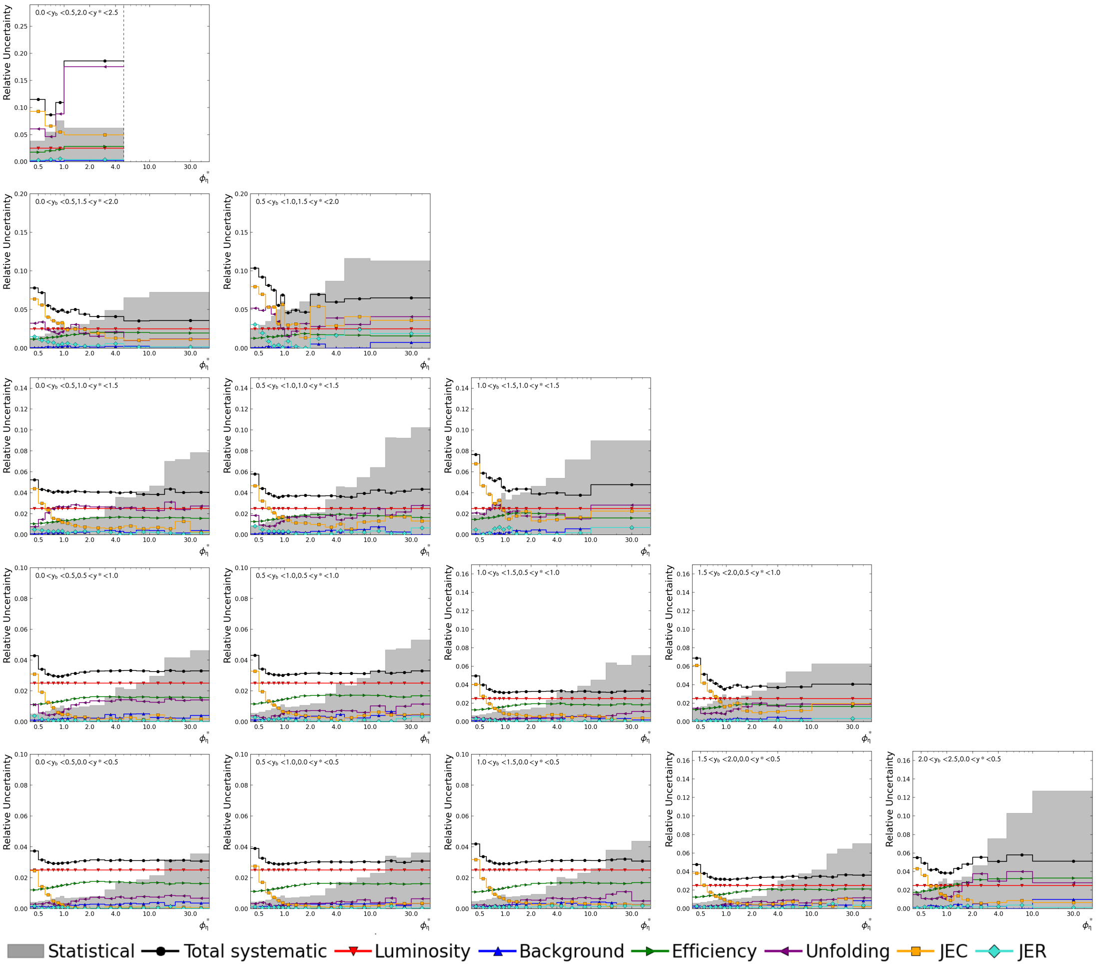


Figure A.10: Total systematic uncertainty (black) and the different uncertainty sources (coloured) as a function of ϕ^*_η in all rapidity bins, compared to the statistical uncertainty (light grey). The uncertainties have been symmetrized by taking half of the spread in each uncertainty source.

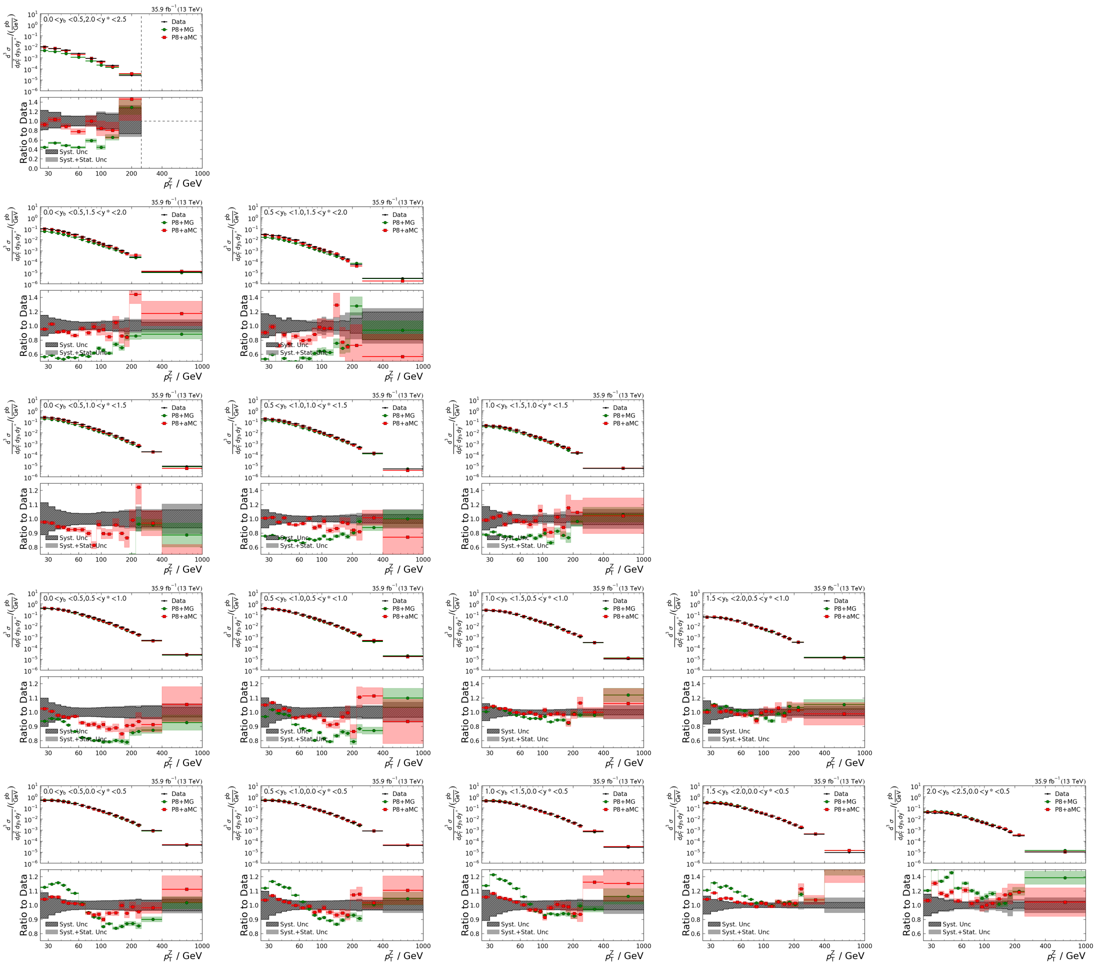


Figure A.11: Comparison of measured cross sections to simulations generated with PYTHIA8 at LO (P8+MG) and NLO (P8+aMC) as a function of $p_{\mathrm{T}}^{\mathrm{Z}}$ in all rapidity bins. In the ratios, the coloured bands represent the statistical uncertainties of the simulations (red and green). The systematic uncertainty of the measurement is shown as hatched grey band, the total experimental uncertainty, i.e. the quadratic sum of systematic and statistical uncertainties, is shown as light grey band.

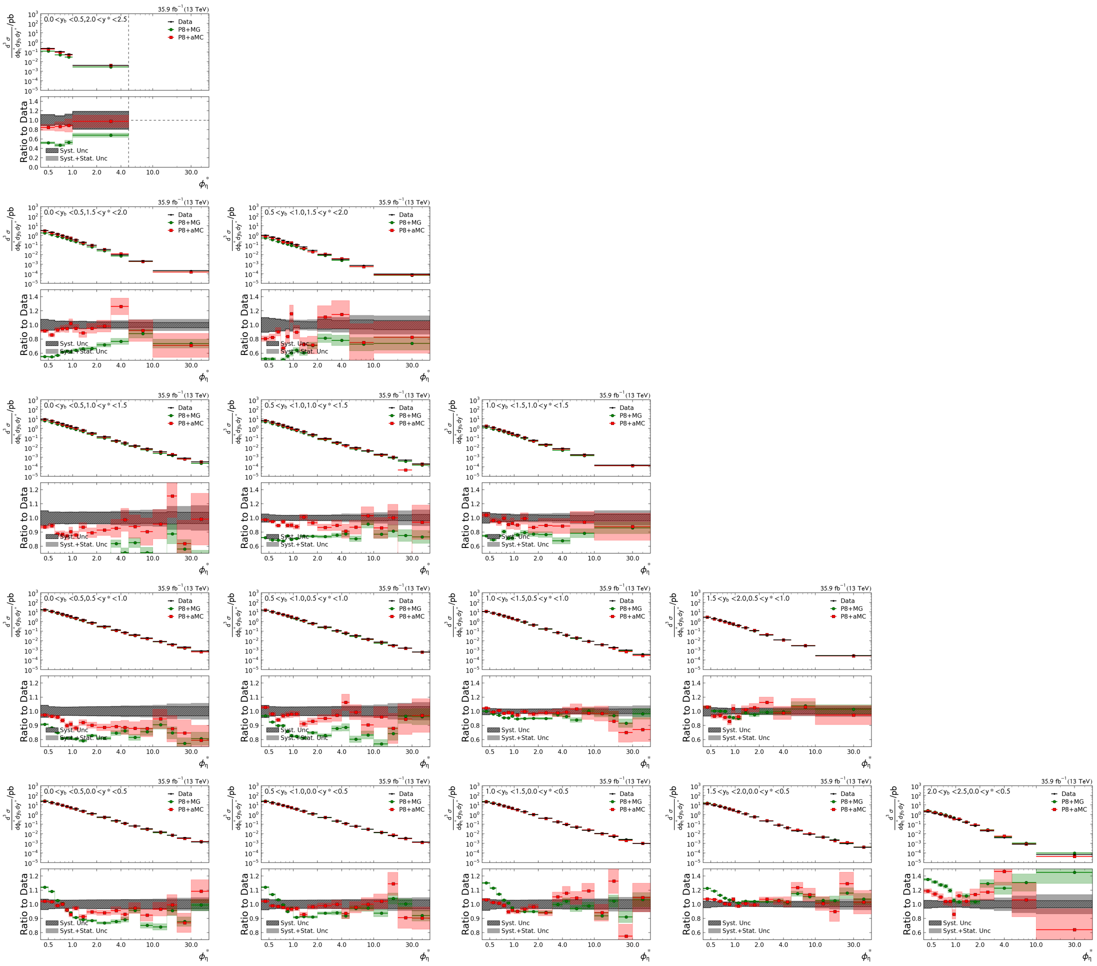


Figure A.12: Comparison of measured cross sections to simulations generated with Pythia8 at LO (P8+MG) and NLO (P8+aMC) as a function of ϕ^*_η in all rapidity bins. In the ratios, the coloured bands represent the statistical uncertainties of the simulations (red and green). The systematic uncertainty of the measurement is shown as hatched grey band, the total experimental uncertainty, i.e. the quadratic sum of systematic and statistical uncertainties, is shown as light grey band.

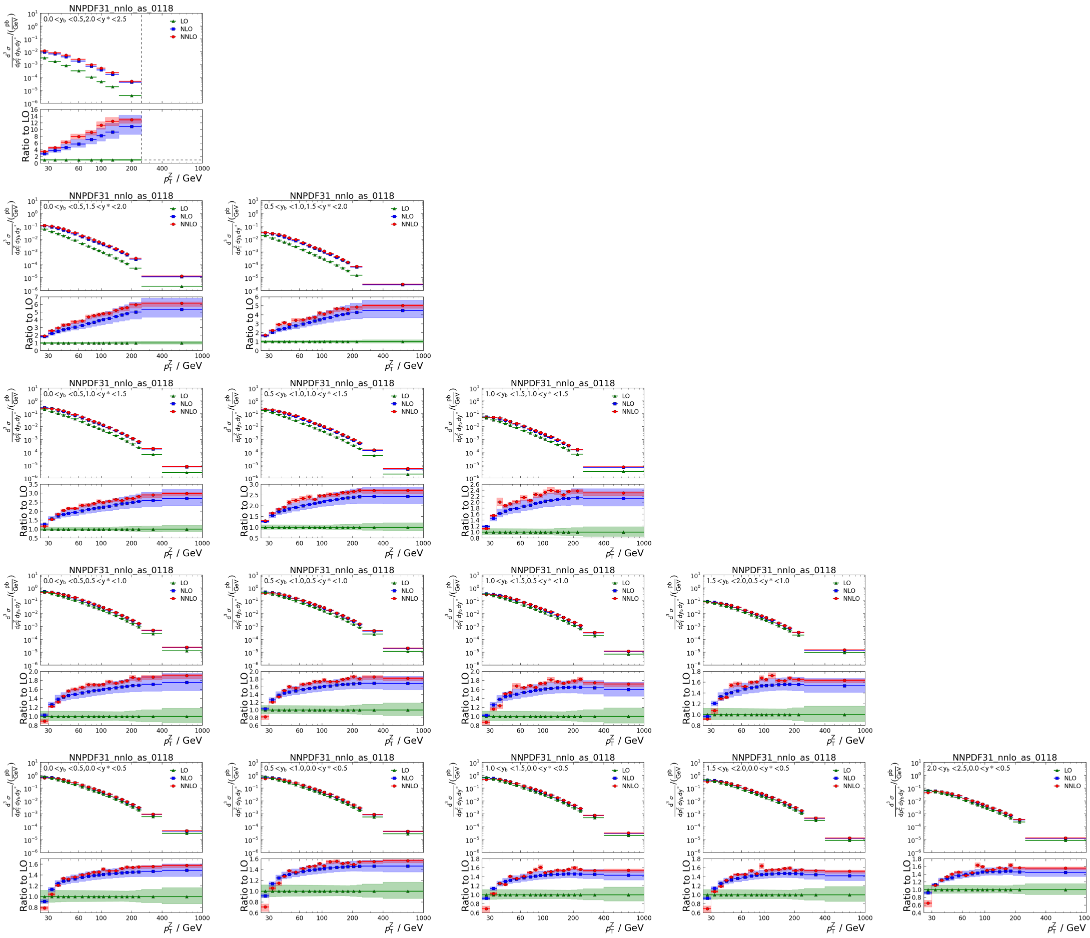


Figure A.13: Comparison of distributions in LO, NLO and NNLO calculations as a function of $p_{\mathrm{T}}^{\mathrm{Z}}$ in all rapidity bins. The coloured bands indicate the scale uncertainty.

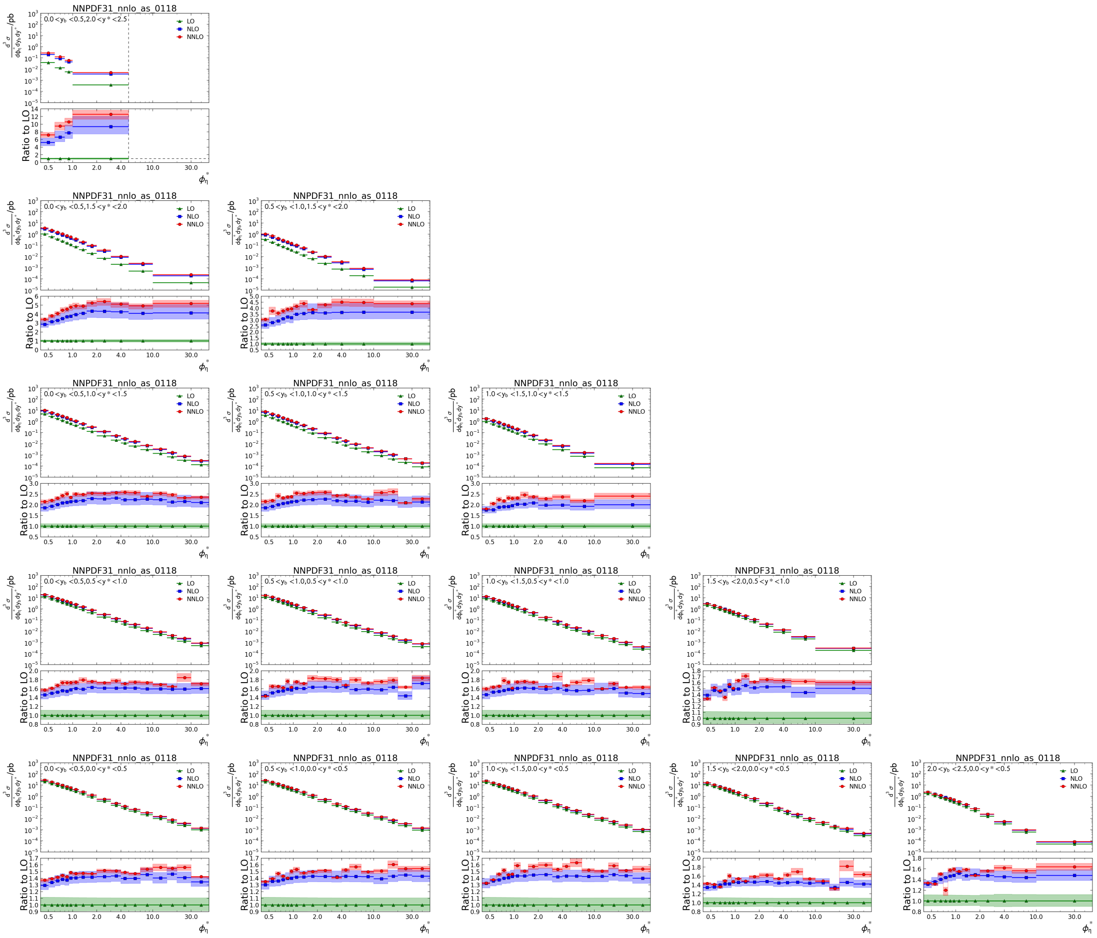


Figure A.14: Comparison of distributions in LO, NLO and NNLO calculations as a function of ϕ^*_η in all rapidity bins. The coloured bands indicate the scale uncertainty.

Figure A.15: Subprocess decomposition of cross section calculations at LO as a function of $p_{\mathrm{T}}^{\mathrm{Z}}$ in all rapidity bins.

Figure A.16: Subprocess decomposition of cross section calculations at LO as a function of ϕ^*_η in all rapidity bins.

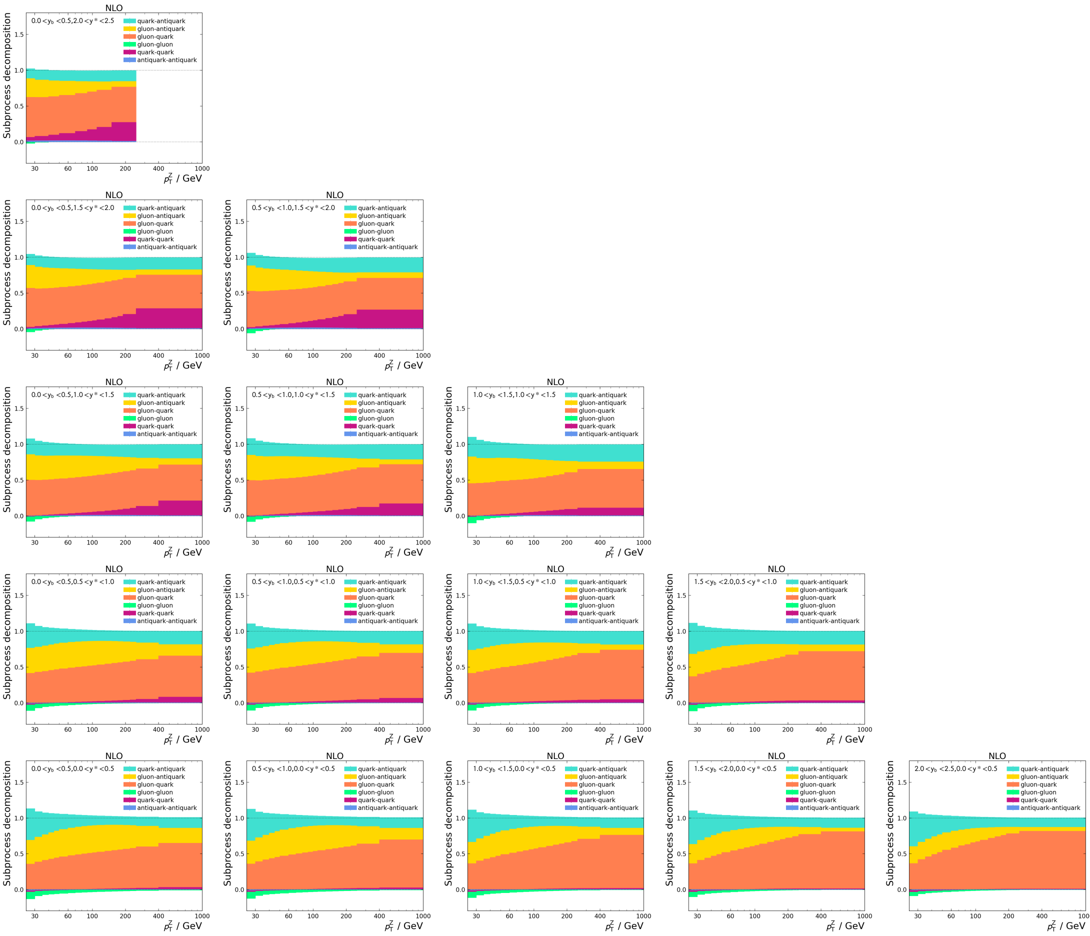


Figure A.17: Subprocess decomposition of cross section calculations at NLO as a function of $p_{\mathrm{T}}^{\mathrm{Z}}$ in all rapidity bins.

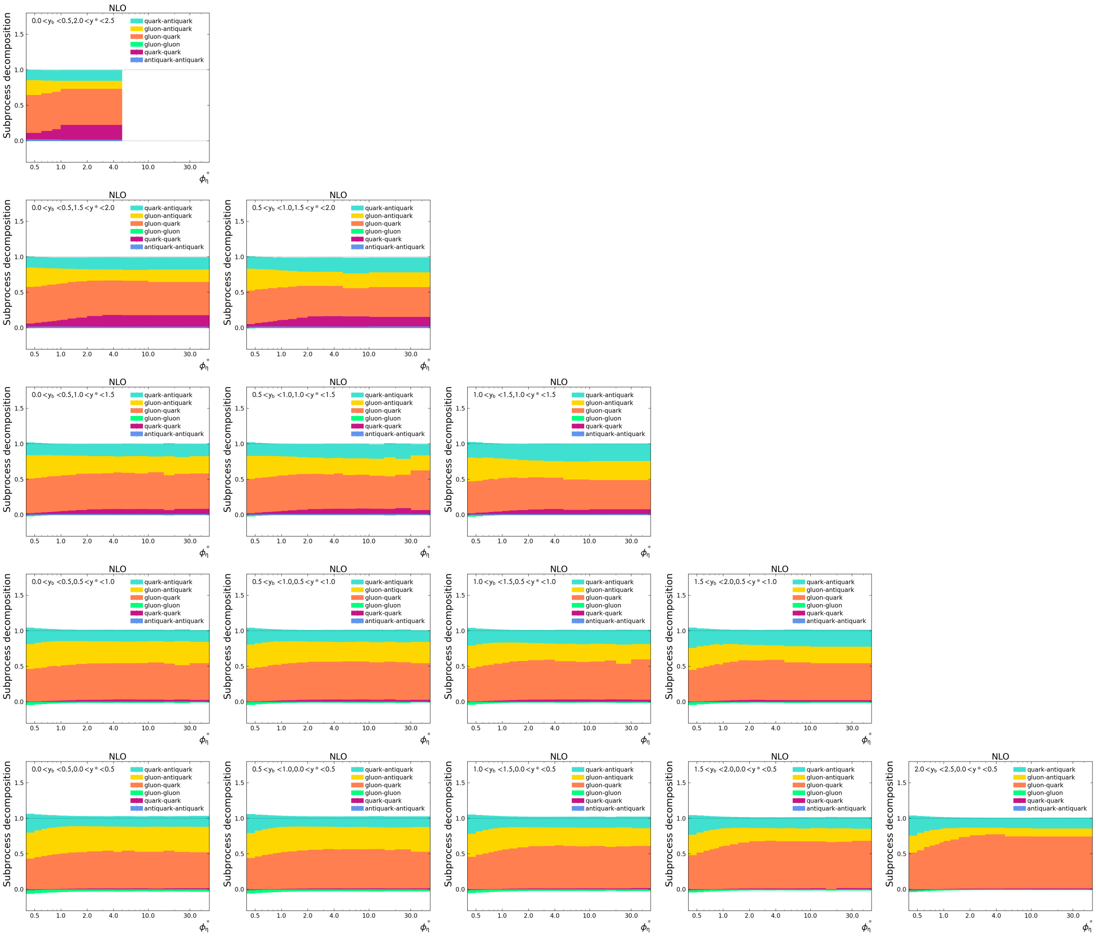


Figure A.18: Subprocess decomposition of cross section calculations at NLO as a function of ϕ^*_η in all rapidity bins.

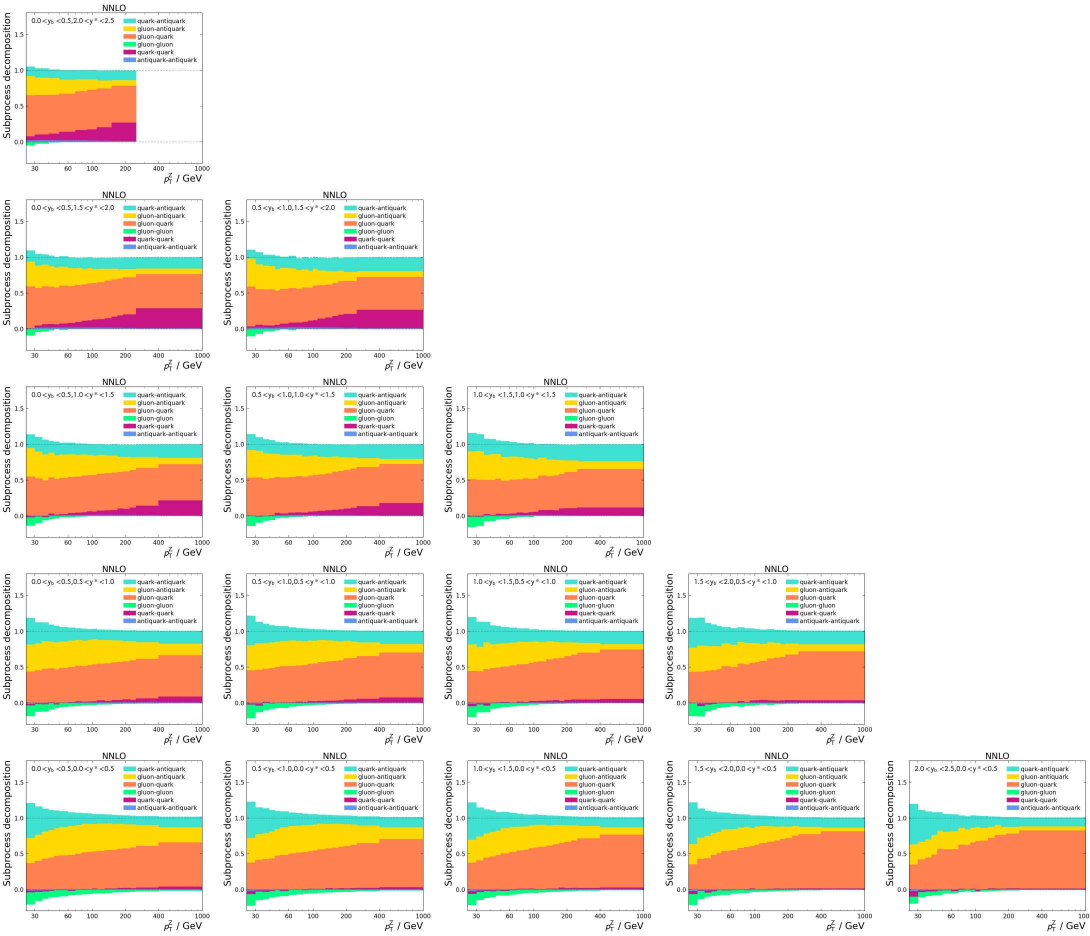


Figure A.19: Subprocess decomposition of cross section calculations at NNLO as a function of $p_{\mathrm{T}}^{\mathrm{Z}}$ in all rapidity bins.

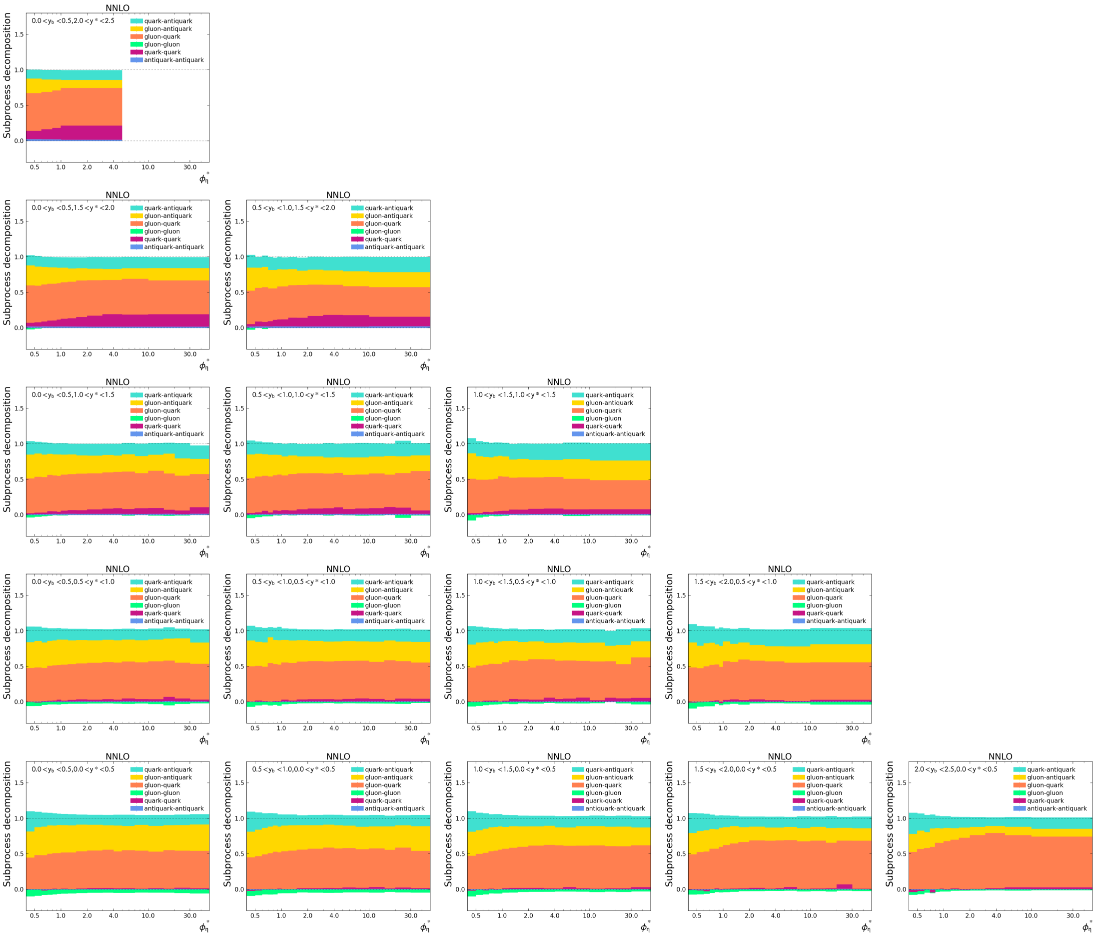


Figure A.20: Subprocess decomposition of cross section calculations at NNLO as a function of ϕ^*_η in all rapidity bins.

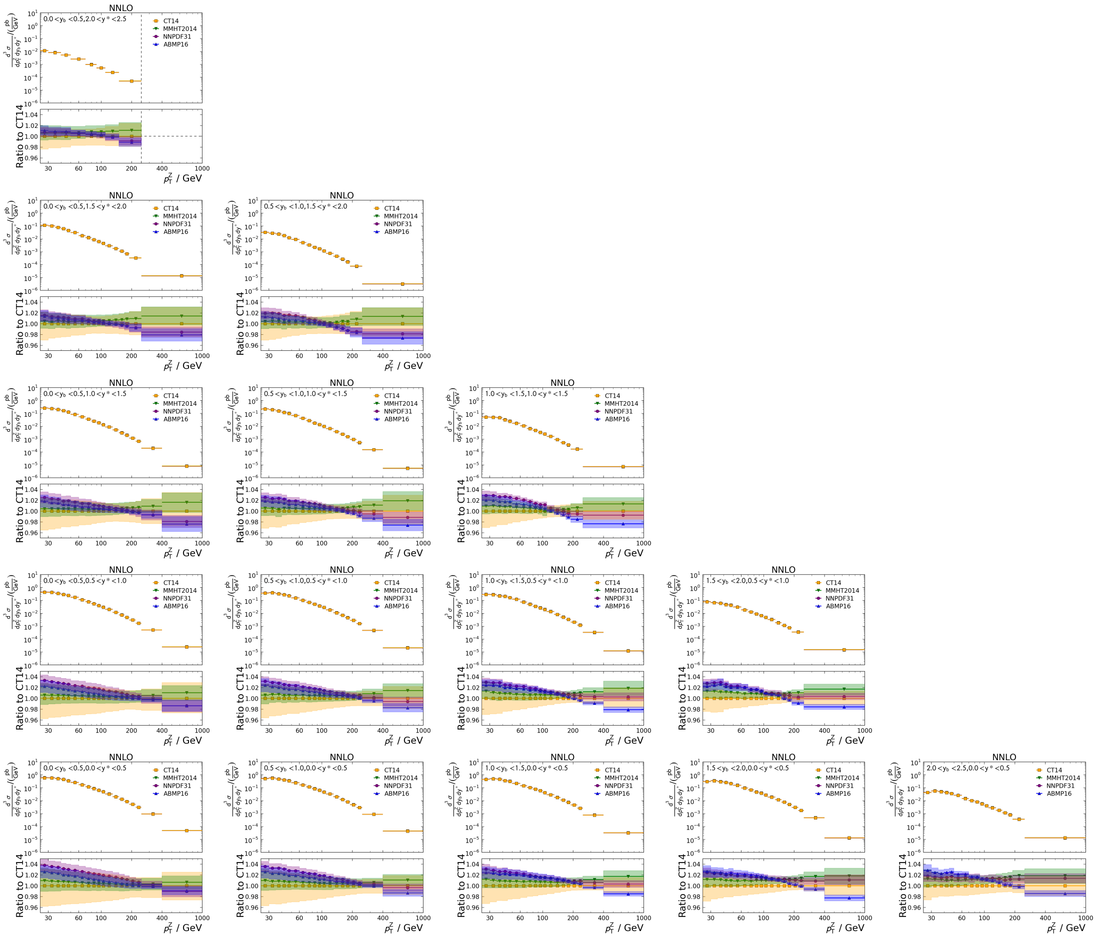


Figure A.21: Comparison of distributions calculated at NNLO for different PDF sets as a function of $p_{\mathrm{T}}^{\mathrm{Z}}$ in all rapidity bins. The coloured bands indicate the PDF uncertainty.

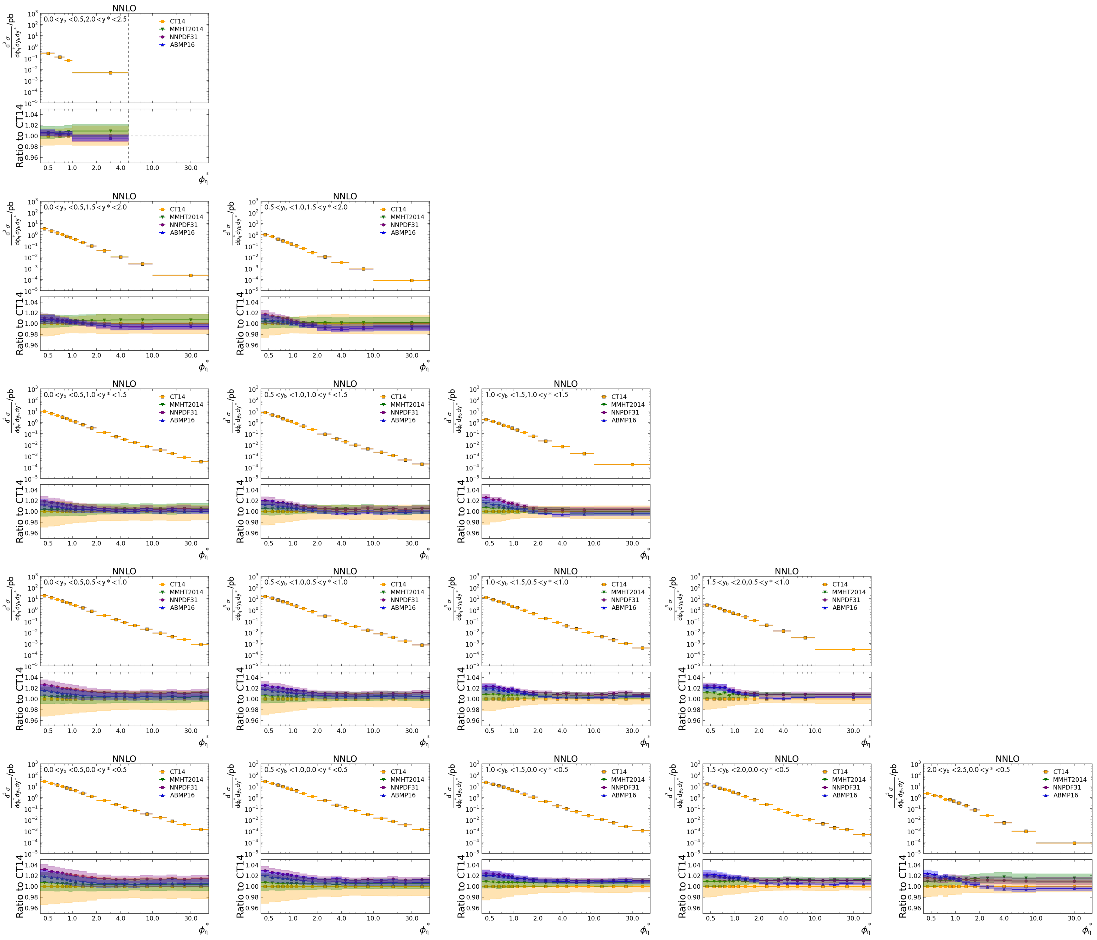


Figure A.22: Comparison of distributions calculated at NNLO for different PDF sets as a function of ϕ^*_η in all rapidity bins. The coloured bands indicate the PDF uncertainty.

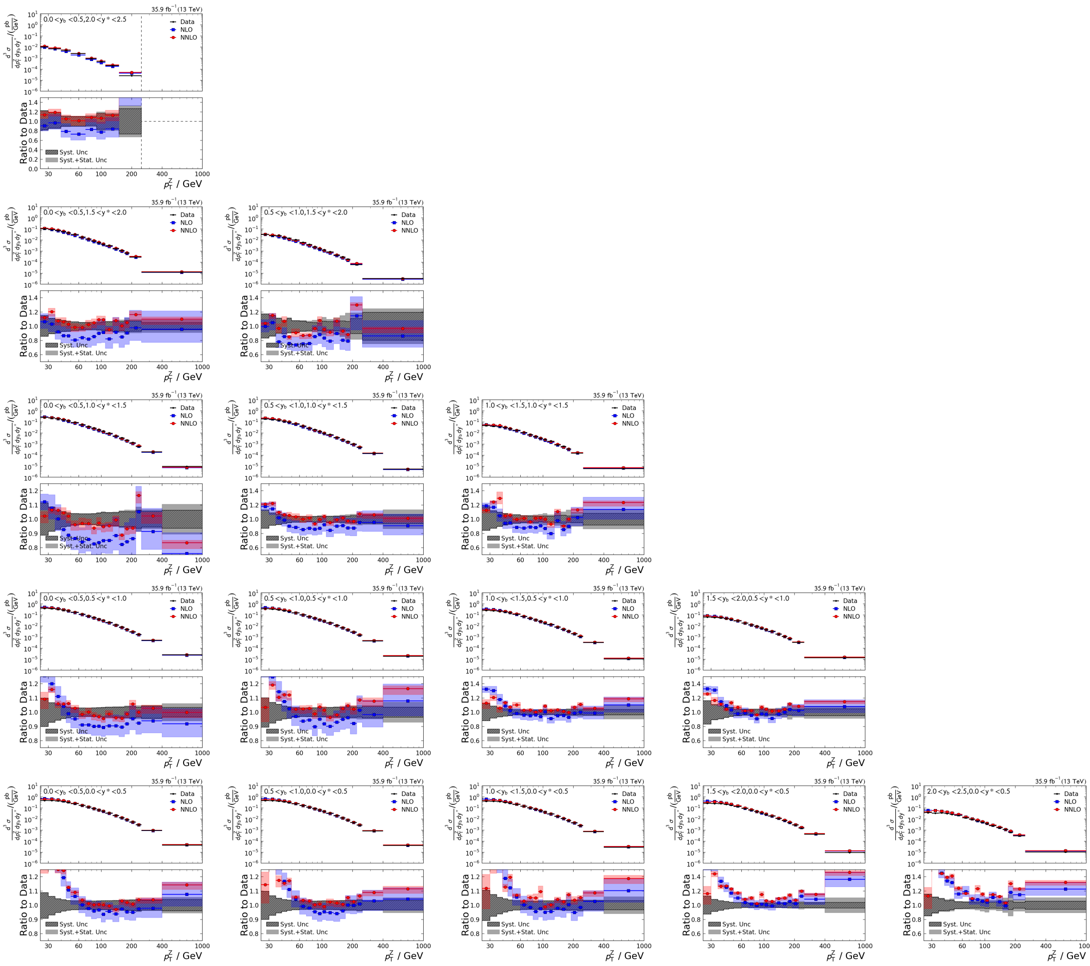


Figure A.23: Comparison of measured cross sections to fixed order calculations at NLO and NNLO precision as a function of $p_{\mathrm{T}}^{\mathrm{Z}}$ in all rapidity bins. In the ratios, the coloured bands represent the theory uncertainties of the calculations (blue and red). The systematic uncertainty of the measurement is shown as hatched grey band, the total experimental uncertainty, i.e. the quadratic sum of systematic and statistical uncertainties, is shown as light grey band.

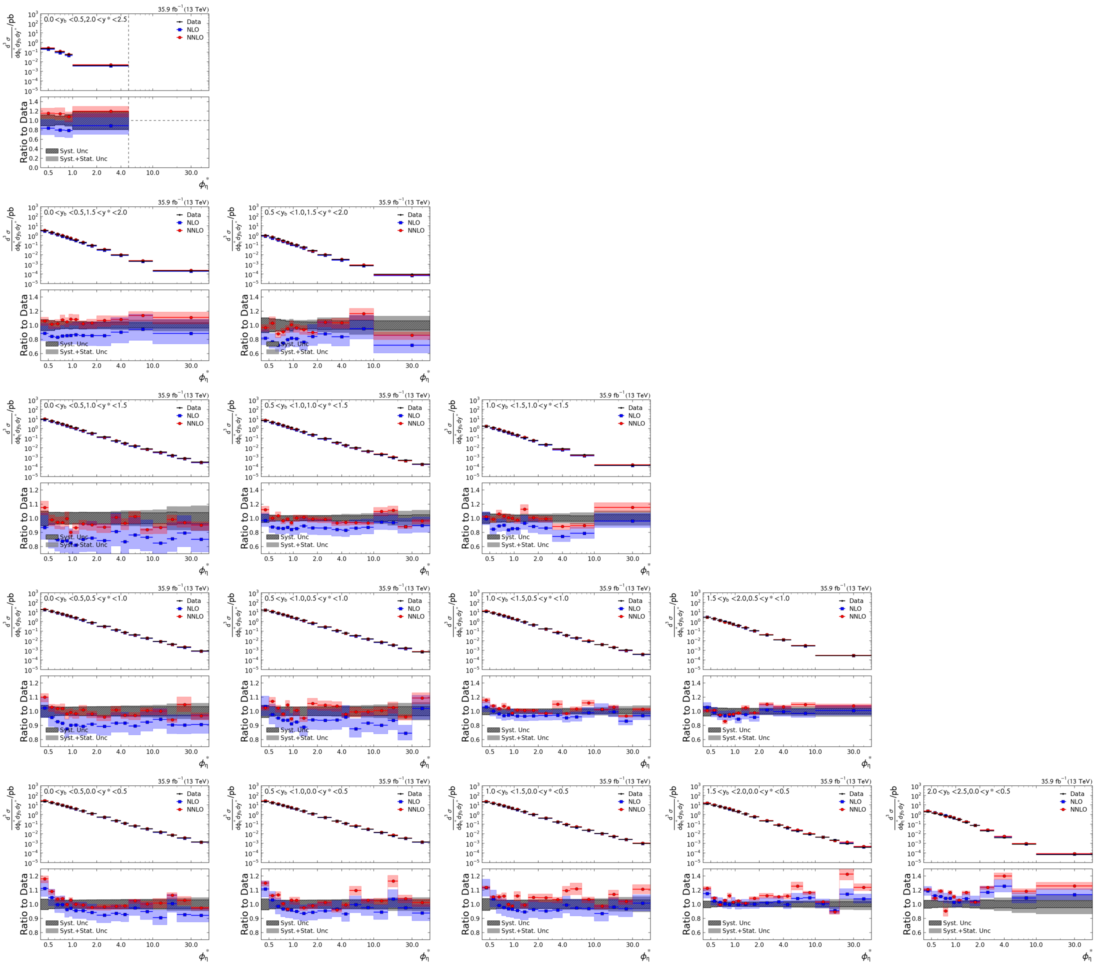


Figure A.24: Comparison of measured cross sections to fixed order calculations at NLO and NNLO precision as a function of ϕ^*_η in all rapidity bins. In the ratios, the coloured bands represent the theory uncertainties of the calculations (blue and red). The systematic uncertainty of the measurement is shown as hatched grey band, the total experimental uncertainty, i.e. the quadratic sum of systematic and statistical uncertainties, is shown as light grey band.

APPENDIX B

List of Figures

APPENDIX C

List of Tables

APPENDIX D

Bibliography

[1] Alison Wright. *Nobel Prize 2013: Englert and Higgs*. Nov. 2013.
DOI: `10.1038/nphys2800`.

[2] ZEUS Collaboration. “PDF Fits at HERA”. *PoS* EPS-HEP2011 (2011), p. 320.
DOI: `10.22323/1.134.0320`. arXiv: `1112.2107 [hep-ph]`.

[3] NNPDF Collaboration. “Parton distributions from high-precision collider data”. *Eur. Phys. J.* C77.10 (2017), p. 663.
DOI: `10.1140/epjc/s10052-017-5199-5`. arXiv: `1706.00428 [hep-ph]`.

[4] Sayipjamal Dulat et al. “New parton distribution functions from a global analysis of quantum chromodynamics”. *Phys. Rev.* D93.3 (2016), p. 033006.
DOI: `10.1103/PhysRevD.93.033006`. arXiv: `1506.07443 [hep-ph]`.

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

[6] S. Alekhin, J. Blumlein, and S. Moch. “The ABM parton distributions tuned to LHC data”. *Phys. Rev.* D89.5 (2014), p. 054028.
DOI: `10.1103/PhysRevD.89.054028`. arXiv: `1310.3059 [hep-ph]`.

[7] Miguel Santos Correa. “Automated Production of Interpolation Grids at NNLO for the Triple-Differential Z+Jet Cross Section Measurement at the LHC”. Masterarbeit. ETP-KA/2018-26. Karlsruhe Institute of Technology, 2018.

[8] Particle Data Group Collaboration. “Review of Particle Physics”. *Phys. Rev.* D98.3 (2018), p. 030001.
DOI: `10.1103/PhysRevD.98.030001`.

[9] CMS Collaboration. “Higgs boson discovery and recent results”. *Proceedings, 22th International Workshop on High Energy Physics and Quantum Field Theory (QFTHEP 2015): Samara, Russia, June 24-July 1, 2015*. 2015. arXiv: `1510.01924 [hep-ex]`.

[10] Peter W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. *Phys. Rev. Lett.* 13 (1964). [,160(1964)], pp. 508–509.
DOI: `10.1103/PhysRevLett.13.508`.

[11] Michael E. Peskin and Daniel V. Schroeder. “An Introduction to quantum field theory”. Reading, USA: Addison-Wesley, 1995. ISBN: 9780201503975, 0201503972.

[12] Yuri L. Dokshitzer. “Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.” *Sov. Phys. JETP* 46 (1977). [Zh. Eksp. Teor. Fiz.73,1216(1977)], pp. 641–653.

[13] V. N. Gribov and L. N. Lipatov. “Deep inelastic e p scattering in perturbation theory”. *Sov. J. Nucl. Phys.* 15 (1972). [Yad. Fiz.15,781(1972)], pp. 438–450.

[14] Guido Altarelli and G. Parisi. “Asymptotic Freedom in Parton Language”. *Nucl. Phys.* B126 (1977), pp. 298–318.
DOI: `10.1016/0550-3213(77)90384-4`.

[15] Daniel Britzger et al. “New features in version 2 of the fastNLO project”. *Proceedings, 20th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2012): Bonn, Germany, March 26-30, 2012*. 2012, pp. 217–221.
DOI: `10.3204/DESY-PROC-2012-02/165`. arXiv: `1208.3641 [hep-ph]`.

[16] Stefan Höche. “Introduction to parton-shower event generators”. *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014): Boulder, Colorado, June 2-27, 2014*. 2015, pp. 235–295.
DOI: `10.1142/9789814678766_0005`. arXiv: `1411.4085 [hep-ph]`.

[17] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. *JHEP* 07 (2014), p. 079.
DOI: `10.1007/JHEP07(2014)079`. arXiv: `1405.0301 [hep-ph]`.

[18] Thomas Gehrmann et al. “Jet cross sections and transverse momentum distributions with NNLOJET”. *PoS* RADCOR2017 (2018), p. 074.
DOI: `10.22323/1.290.0074`. arXiv: `1801.06415 [hep-ph]`.

[19] Torbjörn Sjöstrand. “The PYTHIA Event Generator: Past, Present and Future” (2019). arXiv: `1907.09874 [hep-ph]`.

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

[21] Johannes Bellm et al. “Herwig 7.0/Herwig++ 3.0 release note”. *Eur. Phys. J.* C76.4 (2016), p. 196.
DOI: `10.1140/epjc/s10052-016-4018-8`. arXiv: `1512.01178 [hep-ph]`.

[22] Enrico Bothmann et al. “Event Generation with Sherpa 2.2”. *SciPost Phys.* 7 (2019), p. 034.
DOI: `10.21468/SciPostPhys.7.3.034`. arXiv: `1905.09127 [hep-ph]`.

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

[24] 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 [hep-ph]`.

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

[26] Rikkert Frederix and Stefano Frixione. “Merging meets matching in MC@NLO”. *JHEP* 12 (2012), p. 061.
DOI: `10.1007/JHEP12(2012)061`. arXiv: `1209.6215 [hep-ph]`.

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

[28] Stefano Frixione and Bryan R. Webber. “Matching NLO QCD computations and parton shower simulations”. *JHEP* 06 (2002), p. 029.
DOI: `10.1088/1126-6708/2002/06/029`. arXiv: `hep-ph/0204244 [hep-ph]`.

[29] Bo Andersson et al. “Parton Fragmentation and String Dynamics”. *Phys. Rept.* 97 (1983), pp. 31–145.
DOI: `10.1016/0370-1573(83)90080-7`.

[30] D. Amati and G. Veneziano. “Preconfinement as a Property of Perturbative QCD”. *Phys. Lett.* 83B (1979), pp. 87–92.
DOI: `10.1016/0370-2693(79)90896-7`.

[31] GEANT4 Collaboration. “GEANT4: A Simulation toolkit”. *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303.
DOI: `10.1016/S0168-9002(03)01368-8`.

[32] Geant4 Collaboration. “Recent developments in Geant4”. *Annals Nucl. Energy* 82 (2015), pp. 19–28.
DOI: `10.1016/j.anucene.2014.08.021`.

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

[34] S. D. Ellis et al. “Jets in hadron-hadron collisions”. *Prog. Part. Nucl. Phys.* 60 (2008), pp. 484–551.
DOI: `10.1016/j.ppnp.2007.12.002`. arXiv: `0712.2447 [hep-ph]`.

[35] Klaus Rabbertz. “Jet Physics at the LHC”. *Springer Tracts Mod. Phys.* 268 (2017), pp.1–214.
DOI: `10.1007/978-3-319-42115-5`.

[36] Esma Mobs. “The CERN accelerator complex. Complexe des accélérateurs du CERN”. OPEN-PHO-ACCEL-2016-009 (July 2016). General Photo.

[37] CMS Collaboration Collaboration. “Interactive Slice of the CMS detector”. CMS-OUTREACH-2016-027 (Aug. 2016).

[38] CMS Collaboration. “The CMS Experiment at the CERN LHC”. *JINST* 3 (2008), S08004.
DOI: `10.1088/1748-0221/3/08/S08004`.

[39] CMS Collaboration. “The CMS trigger system”. *JINST* 12.01 (2017), P01020.
DOI: `10.1088/1748-0221/12/01/P01020`. arXiv: `1609.02366 [physics.ins-det]`.

[40] K. Bos et al. “LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)”. Technical Design Report LCG CERN-LHCC-2005-024. Geneva: CERN, 2005.

[41] Konrad Meier et al. “Dynamic provisioning of a HEP computing infrastructure on a shared hybrid HPC system”. *J. Phys. Conf. Ser.* 762.1 (2016), p. 012012.
DOI: `10.1088/1742-6596/762/1/012012`.

[42] CMS Collaboration. “The CMS Particle Flow Algorithm”. *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013): Paris, France, April 22-25, 2013*. 2013, pp. 295–304. arXiv: `1401.8155 [hep-ex]`.

[43] CMS Collaboration Collaboration. “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”. Tech. rep. CMS-PAS-PFT-09-001. Geneva: CERN, Apr. 2009.

[44] CMS Collaboration. “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”. *JINST* 13.06 (2018), P06015.
DOI: `10.1088/1748-0221/13/06/P06015`. arXiv: `1804.04528 [physics.ins-det]`.

[45] CMS Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. *JINST* 9.10 (2014), P10009.
DOI: `10.1088/1748-0221/9/10/P10009`. arXiv: `1405.6569 [physics.ins-det]`.

[46] Matteo Cacciari and Gavin P. Salam. “Pileup subtraction using jet areas”. *Phys. Lett.* B659 (2008), pp. 119–126.
DOI: `10.1016/j.physletb.2007.09.077`. arXiv: `0707.1378 [hep-ph]`.

[47] CMS Collaboration Collaboration. “Performance of the Particle-Flow jet identification criteria using proton-proton collisions at 13 TeV for Run2016 data”. Tech. rep. CMS-AN-17-074. Geneva: CERN, May 2017.

[48] CMS Collaboration Collaboration. “Pileup Jet Identification”. Tech. rep. CMS-PAS-JME-13-005. Geneva: CERN, 2013.

[49] Daniele Bertolini et al. “Pileup Per Particle Identification”. *JHEP* 10 (2014), p. 059.
DOI: `10.1007/JHEP10(2014)059`. arXiv: `1407.6013 [hep-ph]`.

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

[51] CMS Collaboration. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". *JINST* 12.02 (2017), P02014.
DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].

[52] CMS Tracker Group Collaboration. "The construction of the phase-1 upgrade of the CMS pixel detector". *JINST* 12.12 (2017), p. C12049.
DOI: 10.1088/1748-0221/12/12/C12049.

[53] CMS Collaboration. "CMS Luminosity Measurements for the 2016 Data Taking Period". CMS-PAS-LUM-17-001 (2017).

[54] S. D. Drell and Tung-Mow Yan. "Massive Lepton Pair Production in Hadron-Hadron Collisions at High-Energies". *Phys. Rev. Lett.* 25 (1970). [Erratum: Phys. Rev. Lett.25,902(1970)], pp. 316–320.
DOI: 10.1103/PhysRevLett.25.316,10.1103/PhysRevLett.25.902.2.

[55] "Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016". CMS-DP-2018-028 (June 2018).

[56] CMS Collaboration. "Measurements of differential Z boson production cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV" (2019). arXiv: 1909.04133 [hep-ex].

[57] CMS Collaboration. "Measurements of differential production cross sections for a Z boson in association with jets in pp collisions at $\sqrt{s} = 8$ TeV". *JHEP* 04 (2017), p. 022.
DOI: 10.1007/JHEP04(2017)022. arXiv: 1611.03844 [hep-ex].

[58] CMS Collaboration. "Measurement of differential cross sections for Z boson production in association with jets in proton-proton collisions at $\sqrt{s} = 13$ TeV". *Eur. Phys. J.* C78.11 (2018), p. 965.
DOI: 10.1140/epjc/s10052-018-6373-0. arXiv: 1804.05252 [hep-ex].

[59] Dominik Wilhelm Haitz. "Precision Studies of Proton Structure and Jet Energy Scale with the CMS Detector at the LHC". PhD thesis. Karlsruhe U., EKP, 2016.
DOI: 10.5445/IR/1000055680.

[60] Anna Carlotta Friedel. "Measurement of the Z ($\rightarrow \mu\mu$) + jets Cross Section at $\sqrt{s} = 13\,\mathrm{TeV}$ with the CMS Experiment and Studies of the Proton Structure". Masterarbeit. IEKP-KA/2017-15. Karlsruhe Institute of Technology, 2017.

[61] CMS Collaboration. "Measurement of the triple-differential dijet cross section in proton-proton collisions at $\sqrt{s} = 8\,\mathrm{TeV}$ and constraints on parton distribution functions". *Eur. Phys. J.* C77.11 (2017), p. 746.
DOI: 10.1140/epjc/s10052-017-5286-7. arXiv: 1705.02628 [hep-ex].

[62] A. Banfi et al. “Optimisation of variables for studying dilepton transverse momentum distributions at hadron colliders”. *Eur. Phys. J.* C71 (2011), p. 1600. DOI: `10.1140/epjc/s10052-011-1600-y`. arXiv: `1009.1580 [hep-ex]`.

[63] A. Bodek et al. “Extracting Muon Momentum Scale Corrections for Hadron Collider Experiments”. *Eur. Phys. J.* C72 (2012), p. 2194. DOI: `10.1140/epjc/s10052-012-2194-8`. arXiv: `1208.3710 [hep-ex]`.

[64] CMS Collaboration. *Rochester Corrections*. URL: `https://twiki.cern.ch/twiki/bin/view/CMS/RochcorMuon` (visited on 23/09/2019).

[65] CMS Collaboration. *Reference muon id, isolation and trigger efficiencies for Run-II*. URL: `https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffsRun2` (visited on 23/09/2019).

[66] G. D’Agostini. “Improved iterative Bayesian unfolding”. *Alliance Workshop on Unfolding and Data Correction Hamburg, Germany, May 27-28, 2010*. 2010. arXiv: `1010.0632 [physics.data-an]`.

[67] Stefan Schmitt. “TUnfold: an algorithm for correcting migration effects in high energy physics”. *JINST* 7 (2012), T10003. DOI: `10.1088/1748-0221/7/10/T10003`. arXiv: `1205.6201 [physics.data-an]`.

[68] CMS Collaboration. *Recommendations on Unfolding*. URL: `https://twiki.cern.ch/twiki/bin/view/CMS/ScrecUnfolding` (visited on 23/09/2019).

[69] CMS Collaboration. *Recommendations for luminosity estimations*. URL: `https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM` (visited on 23/09/2019).

[70] CMS Collaboration. *Jet Energy Resolution*. URL: `https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution` (visited on 18/11/2019).

[71] Aude Gehrmann-De Ridder et al. “The NNLO QCD corrections to Z boson production at large transverse momentum”. *JHEP* 07 (2016), p. 133. DOI: `10.1007/JHEP07(2016)133`. arXiv: `1605.04295 [hep-ph]`.

[72] Alexander Heidelbach. “Untersuchung eines automatisierten Arbeitsablaufs zur Erstellung von Interpolationsgittern in perturbativer QCD”. Bachelorarbeit. ETP-BACHELOR-KA/2019-15. Karlsruhe Institute of Technology, 2019.

Danksagung

Zunächst gilt mein Dank Prof. Dr. Günter Quast für dessen Engagement, mir den Einstieg in die Teilchenphysik zu ermöglichen und für die Unterstützung während dieser Arbeit. Besonderer Dank gilt Priv.-Doz. Dr. Klaus Rabbertz für seine fachliche Unterstützung, ohne die diese Arbeit nicht möglich gewesen wäre. Zudem bedanke ich mich vielmals für die Gelegenheit, durch meinen Aufenthalt in Genf das CERN leibhaftig in Aktion erleben zu dürfen.

Vielen Dank auch an die komplette Arbeitsgruppe, die mich sehr freundlich aufgenommen hat. Insbesondere dem Computing Team um Christoph Heidecker, Matthias Schnepf, Florian von Cube und auch Frank Fischer vielen Dank für die Unterstützung bei der Beseitigung von störrischen Programmierfehlern. Ebenso vielen Dank an Daniel Savoiu für die zahlreichen Diskussionen, vor allem für jene weit über die Physik hinaus. Zudem vielen Dank an alle, die durch Korrekturlesen zu meiner Arbeit beigetragen haben.

Dank gilt natürlich ebenfalls meiner Familie für die Unterstützung während meiner gesamten Studienzeit. Danke auch an meine jetzigen und auch ehemaligen Mitbewohner, durch die ich neue Freundschaften in Karlsruhe schließen konnte. Und natürlich an alle meine Freunde aus der Studien- und Schulzeit, die wir uns teilweise seit mehr als 20 Jahren die Treue halten und auch weiter halten werden.

Und ganz zuletzt vielen Dank an meine liebe Sabrina, die in den letzten Jahren viel auf sich nehmen musste um mich zu unterstützen!

Erklärung der selbständigen Anfertigung der Dissertationsschrift

Hiermit erkläre ich, dass ich die Dissertationsschrift mit dem Titel

»Jet energy calibration and triple differential
inclusive cross section measurements
with $Z\ (\to \mu\mu)$ *+ jet events at 13 TeV*
recorded by the CMS detector«

selbständig und unter ausschließlicher Verwendung der angegebenen Hilfsmittel angefertigt habe.

Thomas Berger
Karlsruhe, den 09. Oktober 2019